

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number  
**WO 02/08399 A2**

128  
88  
81  
209

(51) International Patent Classification<sup>7</sup>: C12N 9/00

(21) International Application Number: PCT/US01/23092

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/220,038 21 July 2000 (21.07.2000) US  
60/222,112 28 July 2000 (28.07.2000) US  
60/222,831 4 August 2000 (04.08.2000) US  
60/224,729 11 August 2000 (11.08.2000) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(71) Applicant and

(72) Inventor: THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). TRI-BOULEY, Catherine, M. [US/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, MT. View, CA 94043 (US). NGUYEN, Danniell, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). BURFORD, Neil [US/US]; 105 Wildwood Circle, Durham, CT 06422

(US). LAL, Preeti [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). RECIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). GREENWALD, Sara, R. [US/US]; 21 Bucareli Drive, San Francisco, CA 94132 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). GIETZEN, Kimberly, J. [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.



WO 02/08399 A2

## HUMAN KINASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

### BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine

residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

#### Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK



activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Substrates for tyrosine kinases can be identified using anti-phosphotyrosine antibodies to screen tyrosine-phosphorylated cDNA expression libraries. Fish, so named for tyrosine-phosphorylated in Src-transformed fibroblast, is a tyrosine kinase substrate which has been identified by such a technique. Fish has five SH3 domains and a phospho homology (PX) domain. Fish is suggested to be involved in signalling by tyrosine kinases and have a role in the actin cytoskeleton (Lock, P. et al (1998) *EMBO J.* 17:4346-4357).

SHP-2, an SH2-domain-containing phosphotyrosine phosphatase, is a positive signal transducer for several receptor tyrosine kinases (RTKs) and cytokine receptors. Phosphotyrosine phosphatases are critical positive and negative regulators in the intracellular signalling pathways that result in growth-factor-specific cell responses such as mitosis, migration, differentiation, transformation, survival or death. Signal-regulatory proteins (SIRPs) comprise a new gene family of at least 15 members, consisting of two subtypes distinguished by the presence or absence of a cytoplasmic SHP-2-binding domain. The SIRP-alpha subfamily members have a cytoplasmic SHP2-binding domain and includes SIRP-alpha-1, a transmembrane protein, a substrate of activated RTKs and which binds to SH2 domains. SIRPs have a high degree of homology with immune antigen recognition molecules. The SIRP-beta subfamily lacks the cytoplasmic tail. The SIRP-beta-1 gene encodes a polypeptide of 398 amino acids. SIRP family members are generally involved in regulation of signals which define different physiological and pathological processes (Kharitonov, A. et al (1997) *Nature* 386:181-186). Two possible areas of regulation include determination of brain diversity and genetic individuality (Sano, S et al (1999) *Biochem. J.* 344 Pt 3:667-675) and recognition of self which fails in diseases such as hemolytic anemia (Oldenburg, P.-A et al (2000) *Science* 288:2051-2054).

#### Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell

proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the

- 5 cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

- The second messenger dependent protein kinases primarily mediate the effects of second  
10 messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include  
15 thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's  
20 Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

- The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of  
25 eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) *J. Biol. Chem.* 270:14875-14883).

- The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) *J. Biol. Chem.* 273:1357-1364).  
30 The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to

phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

5 The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant  
10 enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey,  
15 P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim,  
20 Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

25 The murine homology to *Caenorhabditis elegans* UNC51, a serine/threonine kinase, has been determined to be required to signal the program of gene expression leading to axon formation from granule cells of the cerebellar cortex (Tomoda, T. et al (1999) Neuron 24:333-346. The human homolog of UNC-51, ULK1, for UNC-51 (*C. elegans*)-like kinase 1, is composed of 1050 amino acids, has a calculated MV of 112.6 kDa and a pI of 8.80. ULK1 has 41% overall sequence similarity to  
30 UNC-51 and is highly conserved among vertebrates including mammals, birds, reptiles, amphibians, and fish. By Northern blot analysis, Kuroyanagi et al have shown ULK1 to be ubiquitously expressed in adult tissues, including skeletal muscle, heart, pancreas, brain, placenta, liver, kidney, and lung while UNC-51 has been specifically located in the nervous system of *C. elegans*. Fish and RH mapping

confirmed the localization of ULK1 to human chromosome 12q24.3. (Kuroyanagi, H. et al (1998) Genomics 51:76-85.

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are 3-kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang, X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1

(IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

MAPKKK6 (MAP3K6) is one of numerous MAP3Ks identified. Isolated from skeletal muscle, MAP3K6 is 1,280 amino acids in length with 11 kinase subdomains and is detected in several tissues. The highest expression has been found in heart and skeletal muscle. MAP3K6 has 45% amino acid sequence identity with MAP3K5, while their catalytic domains share 82% identity. MAP3K6 interaction with MAP3K5 *in vivo* was confirmed by coimmunoprecipitation. Recombinant MAP3K6 has been shown to weakly activate the JNK but not the p38 kinase or ERK pathways (Wang, X.S. et al. *supra*)

#### Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

#### Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is

to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

#### Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

#### 5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol. Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRK1 and DRK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced

apoptosis (Sanjo et al., supra). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., supra). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., supra).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate

dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

## 5 KINASES WITH NON-PROTEIN SUBSTRATES

### Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member  
 10 phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma  
 15 membrane by inositol kinases, thus converting PI residues to the biphosphate state ( $PIP_2$ ).  $PIP_2$  is then cleaved into inositol triphosphate ( $IP_3$ ) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

20 PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as  
 25 the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate ( $PIP_2$ ) to PI (3,4,5)  $P_3$  ( $PIP_3$ ).  $PIP_3$  then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such  
 30 as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, supra).

An example of lipid kinase phosphorylation activity is the phosphorylation of



D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP  
5 regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive  
10 inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

#### Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These  
15 two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) *Cancer Res.* 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of  
20 ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zelevnikar, R.J. et al. (1995) *J. Biol. Chem.* 270:7311-7319). Thus AdK may have a pivotal  
25 role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

30 GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second

messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis.

- 5 Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and bucciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. 10 (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

#### Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine 15 kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. 20 Ther. 67:155-186).

The discovery of new human kinases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of 25 nucleic acid and amino acid sequences of human kinases.

### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," 30 "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," "PKIN-12," "PKIN-13," "PKIN-14," "PKIN-15," "PKIN-16," "PKIN-17," "PKIN-18," "PKIN-19," and "PKIN-20." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a

polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence  
5 selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence  
10 at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-20. In  
15 another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:21-40.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least  
20 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the  
25 invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least  
30 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is

transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID

NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an

amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test

compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each

polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

5        Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

10       Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

15       Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

20       It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

25       Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the  
30       cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.



## DEFINITIONS

"PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

- 5       The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

- 10       An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times
- 15       in a given sequence.

- "Altered" nucleic acid sequences encoding PKIN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of
- 20       the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PKIN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,
- 25       hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains
- 30       having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

      The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.

- 5 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by  
10 directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments  
15 containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

20 The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used  
25 to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified  
30 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring

nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys

	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
5	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
10	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
15	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the

evolution of new protein functions.

A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a  
5 fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected  
10 from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:21-40 comprises a region of unique polynucleotide sequence that  
15 specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:21-40 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:21-40 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely  
20 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-20 is encoded by a fragment of SEQ ID NO:21-40. A fragment of SEQ ID NO:1-20 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-20. For example, a fragment of SEQ ID NO:1-20 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-20.  
25 The precise length of a fragment of SEQ ID NO:1-20 and the region of SEQ ID NO:1-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full  
30 length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to

the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

5           Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS  
10 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms  
15 is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other  
20 polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to  
25 compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

30 *Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Open Gap: 11 and Extension Gap: 1 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

10       “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

15       “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.

20       Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989)



Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. 5 Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, 10 such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid 15 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

20 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 25 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

30 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, 5 polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably 10 linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of 15 amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the 20 art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

"Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. 25 Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

30 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers

may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence

that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have  
5 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is  
10 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

15 "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear  
20 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell,  
25 chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure  
30 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

5 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,  
10 trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods  
15 well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as  
20 an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The  
25 nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria,  
30 fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human human kinases (PKIN), the polynucleotides encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a

single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human kinases. For example, SEQ ID NO:2 is 97% identical to mouse tousled-like kinase (GenBank ID g2853031) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:2 is a tousled-like kinase. In an alternative example, SEQ ID NO:10 is 63% identical to human serine/threonine protein kinase (GenBank ID g36615) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The

BLAST probability score is  $7.7e-122$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a serine/threonine kinase. Note that "serine/threonine kinase" is a specific class of kinases. In an alternative example, SEQ ID NO:16 is 53% identical to human receptor protein-tyrosine kinase (GenBank ID g551608) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $4.1e-290$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a receptor tyrosine kinase. In an alternative example, SEQ ID NO:19 is 93% identical to rat Calcium/calmodulin-dependent protein kinase isoform IV (GenBank ID g1836161) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $6.0e-257$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a protein kinase. SEQ ID NO:1, SEQ ID NO:3-9, SEQ ID NO:11-15, SEQ ID NO:17-18, and SEQ ID NO:20 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-20 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:21-40 or that distinguish between SEQ ID NO:21-40 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA



sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2564295H1 is the identification number of an Incyte cDNA sequence, and ADRETUT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71191190V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g1164223) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from

genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:21-40, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-

511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National

Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been  
5 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze  
10 the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire  
15 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or  
20 fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but  
25 not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

30 The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve

the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals

include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted,

5 exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

10 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus);  
20 plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola,  
25 M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1  
5 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol.  
10 Chem. 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH  
15 promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

20 Plant systems may also be used for expression of PKIN. Transcription of sequences encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Bröglie, R. et al.  
25 (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases  
30 where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.



Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest

is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing  
5 polynucleotides which encode PKIN may be designed to contain signal sequences which direct secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation,  
10 lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and  
15 processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may  
20 facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion  
25 proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN  
30 may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in

vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

5 PKIN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of  
10 PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for  
15 these compounds involves producing appropriate cells which express PKIN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

20 An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.  
25 Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN  
30 activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in

vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

5 In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse  
10 embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner  
15 (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential  
20 therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al.  
25 (1998) Science 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae  
30 are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with bladder cancer, prostatic, ovarian, brain, colon, ileum, penis, skin, adrenal tumor, digestive, and cancerous tissues. Therefore, PKIN appears to play a role in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to increase the expression or activity of PKIN.

Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy,

adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, 5 primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased 10 expression or activity of PKIN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

15 In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such 20 disorders include, but are not limited to, those cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

25 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate 30 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic



efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also  
5 be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and  
10 others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG  
15 (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN  
20 amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma  
25 technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate  
30 antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single

chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell,

J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

- (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.
- 15 In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and
- 20 (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA),

25 PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA

30 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

and Blau, H.M. *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to

be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) 5 Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be 10 especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. 15 Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, 20 ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to 25 those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During 30 alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus

genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques

for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs  
5 that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages  
10 within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a  
15 compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective  
20 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the  
25 polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary  
30 library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample



may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,

intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are

used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

5       The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response  
10 to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

15 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the  
20 diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and  
25 may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal or standard values for PKIN expression are established by combining body fluids or cell extracts taken  
30 from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation

between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect  
5 and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to  
10 identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

15 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the  
20 cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels,  
25 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain,  
30 breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress

syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,

5 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,

10 thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and

15 cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker

20 muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly,

25 craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris,

30 myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy,

myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary  
10 hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid  
15 storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with  
20 hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick,  
25 pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample  
30 from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample

indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease

in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic



profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number  
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
15 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention  
20 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000)  
25 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested  
30 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not

necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed  
5 gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be  
10 quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
15 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
20 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
25 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass  
30 spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA*

94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation,

inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

In additional embodiments, the nucleotide sequences which encode PKIN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/220,038, U.S. Ser. No. 60/222,112, U.S. Ser. No. 60/222,831, and U.S. Ser. No. 60/224,729 are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD

database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl  
5 cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles  
10 (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the  
15 UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-  
20 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto  
25 CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo  
30 excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP

96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative human kinases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an



assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode human kinases, the encoded polypeptides were analyzed by querying against

5 PFAM models for human kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as human kinases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as

10 extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly

15 process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

##### **"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene

20 identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a

25 full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals

30 thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent

type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### 5 **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST  
10 analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous  
15 genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:21-40 were compared with  
20 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:21-40 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for  
25 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between  
30 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid

markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## 5 VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

10 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

15

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the  
 20 length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by  
 25 gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced, either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the  
 30 other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are

assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### **VIII. Extension of PKIN Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4

repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs,

genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of 5 [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase  
10 I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.  
15 Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical  
20 microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may  
25 be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may  
30 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### **Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C

oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,



although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

#### **XII. Expression of PKIN**

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D-thiogalactopyranoside

5 (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter  
10 drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-  
15 transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at  
20 specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in  
25 Examples XVI, XVII, XVIII, and XIX where applicable.

### XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice  
30 include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a

marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XIV. Production of PKIN Specific Antibodies**

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-

Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate,  
5 blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring PKIN Using Specific Antibodies**

Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by  
10 covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength  
15 buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

#### **XVI. Identification of Molecules Which Interact with PKIN**

PKIN, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent.  
20 (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

25 Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
30 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVII. Demonstration of PKIN Activity**

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein

substrate by PKIN in the presence of [ $\gamma$ - $^{32}\text{P}$ ]ATP. PKIN is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PKIN. A determination of the  
5 specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate  
10 and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase  
15 sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34<sup>cdc2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and  
20 ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated in an assay containing PKIN, 50 $\mu\text{l}$  of kinase buffer, 1 $\mu\text{g}$  substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu\text{g}$  ATP, and 0.5  $\mu\text{Ci}$  [ $\gamma$ - $^{32}\text{P}$ ]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ - $^{32}\text{P}$ ]ATP is removed by  
25 washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PKIN.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by  
30 the incorporation of  $^{32}\text{P}$  from [ $\gamma$ - $^{32}\text{P}$ ]ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and  $^{32}\text{P}$ -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected

to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA),  
5 scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

### **XVIII. Enhancement/Inhibition of Protein Kinase Activity**

Agonists or antagonists of PKIN activation or inhibition may be tested using assays described  
10 in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

### **XIX. Kinase Binding Assay**

Binding of PKIN to a FLAG-CD44 cyt fusion protein can be determined by incubating PKIN  
to anti-PKIN-conjugated immunoaffinity beads followed by incubating portions of the beads (having  
15 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1 % bovine serum albumin, and 0.05% Triton X-100) in the presence of <sup>125</sup>I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein ) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated <sup>32</sup>P is proportional  
20 to the amount of bound PKIN.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be  
25 understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2564295	1	2564295CD1	21	2564295CB1
2837050	2	2837050CD1	22	2837050CB1
7474590	3	7474590CD1	23	7474590CB1
7474594	4	7474594CD1	24	7474594CB1
7477585	5	7477585CD1	25	7477585CB1
7477587	6	7477587CD1	26	7477587CB1
7594537	7	7594537CD1	27	7594537CB1
70467491	8	70467491CD1	28	70467491CB1
7478559	9	7478559CD1	29	7478559CB1
1698381	10	1698381CD1	30	1698381CB1
7474637	11	7474637CD1	31	7474637CB1
7170260	12	7170260CD1	32	7170260CB1
1797506	13	1797506CD1	33	1797506CB1
1851973	14	1851973CD1	34	1851973CB1
7474604	15	7474604CD1	35	7474604CB1
7474721	16	7474721CD1	36	7474721CB1
7478815	17	7478815CD1	37	7478815CB1
7477141	18	7477141CD1	38	7477141CB1
2190612	19	2190612CD1	39	2190612CB1
7477549	20	7477549CD1	40	7477549CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	2564295CD1	g186555	0.0	Insulin receptor-related receptor [Homo sapiens]
2	2837050CD1	g2853031	0.0	Tousled-like kinase [Mus musculus]
3	7474590CD1	g6453611	5.1e-86	Protein kinase (mutant form) [Mus musculus]
4	7474594CD1	g3879221	5.6e-99	Predicted using Genefinder similar to casein kinase I [Caenorhabditis elegans]
5	7477585CD1	g348245	3.5e-62	Protein serine/threonine kinase [Homo sapiens]
6	7477587CD1	g312998	7.4e-73	Protein kinase [Homo sapiens]
7	7594537CD1	g485398	0.0	90kDa-diacylglycerol kinase [Rattus norvegicus]
8	70467491CD1	g3089349	0.0	Cdc25C associated protein kinase C- TAK1 [Homo sapiens]
9	7478559CD1	g7960111 g9998952	4.2e-114 1.00E-123	Ethanolamine kinase [Homo sapiens] [Homo sapiens] ethanolamine kinase Lykidis, A. et al. Overexpression of a mammalian ethanolamine-specific kinase accelerates the CDP-ethanolamine pathway J. Biol. Chem. 276, 2174-2179 (2001)
10	1698381CD1	g36615	7.7e-122	[Homo sapiens] serine/threonine protein kinase Meyerson, M. et al. (1992) EMBO J. 11:2909-2917
11	7474637CD1	g1181079	0.0	[Homo sapiens] diacylglycerol kinase delta Sakane, F. et al. (1996) J. Biol. Chem. 271:8394-8401
		g1401232	0	[Cricetinae gen. sp.] diacylglycerol kinase eta Klauck, T.M. et al. Cloning and characterization of a glucocorticoid-induced diacylglycerol kinase J. Biol. Chem. 271, 19781-19788 (1996)



Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
12	7170260CD1	g8101585	3.5e-126	[Mus musculus] testis specific serine kinase-3 Zuercher, G. et al (2000) Mech. Dev. 93:175-177
13	1797506CD1	g3300094	4.5e-227	[Homo sapiens] protein kinase/endoribonuclease Tirasophon, W. et al. (1998) Genes Dev. 12:1812-1824
		g12407081	0	[Homo sapiens] protein kinase/ribonuclease IRE1 beta Iwawaki, T. et al. Translational control by the ER transmembrane kinase/ribonuclease IRE1 under ER stress Nat. Cell Biol. 3, 158-164 (2001)
14	1851973CD1	g1853976	1.3e-37	[Schizosaccharomyces pombe] protein kinase Samejima, I., and Yanagida, M. (1994) Mol Cell Biol 14:6361-71
		g9294489	5.00E-47	[Arabidopsis thaliana] IRE homolog; protein kinase-like protein Sato, S., Nakamura, Y., Kaneko, T., Kato, T. et al. Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones DNA Res. 7, 131-135 (2000)
15	7474604CD1	g1171250	2.0e-218	[Mus musculus] protein kinase related to Raf protein kinases Therrien, M. et al. (1995) Cell 83:879-888
16	7474721CD1	g551608	4.1e-290	[Homo sapiens] receptor protein-tyrosine kinase Fox, G.M. et al. (1995) Oncogene 10:897-905
17	7478815CD1	g2873349	0.0	[Homo sapiens] Hexokinase I Ruzzo, A. et al. (1998) Biochem. J. 331(Pt 2):607-613

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
18	7477141CD1	g7239696	1.6e-87	[Homo sapiens] myosin light chain kinase Garcia, J.G. et al. (1997) Am. J. Respir. Cell Mol. Biol. 16:489-494 Garcia, J.G.N. et al. (1996) Biochem. Biophys. Res. Commun. 1:1-1
		g11385416	0	[Mus musculus] striated muscle-specific serine/threonine protein kinase Hsieh, C.M. et al. Striated Muscle Preferentially Expressed Genes alpha and beta Are Two Serine/Threonine Protein Kinases Derived from the Same Gene as the Aortic Preferentially Expressed Gene-1 J. Biol. Chem. 275 (47), 36966-36973 (2000)
19	2190612CD1	g1836161	6.0e-257	[Rattus sp.] Ca2+/calmodulin-dependent protein kinase IV kinase Okuno, S., Kitani, T. and Fujisawa, H. (1996) J. Biochem. 119:1176-1181
20	7477549CD1	g5006445	3.6e-179	[Homo sapiens] CDC42-binding protein kinase beta Moncrieff, C.L. et al. (1999) Genomics 57:297-300
		g2736151	0	[Rattus norvegicus] myotonic dystrophy kinase-related Cdc42-binding kinase Leung, T. et al. Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization Mol. Cell. Biol. 18, 130-140 (1998)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
20		g2217968	1.40E-161	[Homo sapiens] myotonic dystrophy protein kinase Kedra, D. et al. The germinal center kinase gene and a novel CDC25-like gene are located in the vicinity of the PYGM gene on 11q13 Hum. Genet. 100, 611-619 (1997)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2564295CD1	1297	S151 S238 S271 S49 S564 S666 S741 S758 S827 S887 S900 S93 S962 T223 T348 T475 T486 T494 T581 T582 T629 T64 Y454 Y652 T1014 T1020 T1063 S1163 T1171 T1187 S1245 T1275 T1284 S1073 T1128 S1253 T1145	N311 N411 N47 N492 N528 N616 N634 N756 N885 N898 N949	PROTEIN KINASE DOMAIN DM00004   P14617   980-1238: S980-F1239 RECEPTOR PRECURSOR SIGNAL TRANSFERASE TYROSINEPROTEIN KINASE TRANSMEMBRANE GLYCOPROTEIN ATPBINDING PHOSPHORYLATION PD006834: A603-R745, F760-I818 RECEPTOR PRECURSOR SIGNAL TRANSFERASE TYROSINEPROTEIN KINASE TRANSMEMBRANE GLYCOPROTEIN ATPBINDING PHOSPHORYLATION PD005347: Q466-P602 PUTATIVE INSULINLIKE PEPTIDE RECEPTOR PRECURSOR EC 2.7.1.112 TRANSFERASE TYROSINEPROTEIN KINASE TRANSMEMBRANE GLYCOPROTEIN ATPBINDING PHOSPHORYLATION SIGNAL PD146134: L344-E495, V773-G899, D513-C799, E825-R855 PRECURSOR SIGNAL INSULINLIKE RECEPTOR TRANSFERASE TYROSINEPROTEIN KINASE TRANSMEMBRANE GLYCOPROTEIN ATPBINDING PD004354: V330-G410 Receptor tyrosine kinase BL00239: G464-P473, E1030-E1077, M1092-R1114, A1117-E1142, D1144-Y1193, N1198-I1242 Receptor tyrosine kinase BL00240F: T1143-E1190 Receptor tyrosine kinase BL00790H: S831-L856 Tyrosine kinase catalytic domain PR00109: M1059-R1072, Y1105-V1123, L1154-L1164, S1173-G1195, C1217-F1239 Protein kinases signatures and profile protein_kinase tyr.pr: E1091-T1143 Receptor tyrosine kinase class II signature receptor_tyr kin ii.pr: R1119-G1167 Signal peptide: M1-D25 Transmembrane domain: V922-Y944, Furin-like cysteine rich region: G173-K329 Receptor L domain: N47-N170, G346-N472 Eukaryotic protein kinase domain pkinase: I979-E1248 Protein_Kinase_Atp L985-K1013	BLAST-DOMO  BLAST-PRODOM  BLAST-PRODOM  BLAST-PRODOM  BLAST-PRODOM  BLIMPS-BLOCKS  BLIMPS-BLOCKS  BLIMPS-BLOCKS  BLIMPS-PRINTS  PROFILES-SCAN PROFILES-SCAN  HMMER HMMER HMMER-PFAM HMMER-PFAM HMMER-PFAM  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	2837050CD1	718	S165 S186 S194 S238 S246 S257 S275 S298 S44 S46 S509 S605 S632 T176 T269 T344 T403 T488 T558 T78 Y571 Y97	N340 N36 N548 N630 N713 N714	PROTEIN KINASE DOMAIN DM00004 P34314 736-1002: L409-D677 TOUSLEDLIKE KINASE PD102959: M2-E183 KINASE PROTEIN TOUSLEDLIKE PD013350: M237-D400, Q287-L409 TOUSLEDLIKE KINASE KIAA0137 PROTEIN PD035377: K184-T236 TOUSLEDLIKE KINASE MULTIPLE TESTIS TRANSCRIPT PD026280: A682-N718 Tyrosine kinase catalytic domain PR00109: L490-K503, V608-N630 Protein kinases signatures and profile protein_kinase_tyr.prf: E512-S570 Eukaryotic protein kinase domain pkinase: Y408-L687 Protein_Kinase_Atp: L414-K437 Protein_Kinase_St: I534-L546 PROTEIN KINASE DOMAIN DM00004 P27448 58-297: V30-T265 Tyrosine kinase catalytic domain PR00109: Y136-V154, V202-S224, L244-A266 Protein kinases signatures and profile protein_kinase_tyr.prf: Q94-G174 Eukaryotic protein kinase domain pkinase: Y28-L275 Protein_Kinase_St: V142-V154 PROTEIN KINASE DOMAIN DM00004 P48730 11-265: K144-Y392 SIMILAR TO CASEIN KINASES PD115501: F332-D422, L130-T233 Eukaryotic protein kinase domain pkinase: W140-F374 Protein_Kinase_Atp: I146-K169 Signal cleavage: M1-L19	BLAST-DOMO BLAST-PRODUM BLAST-PRODUM BLAST-PRODUM BLAST-PRODUM BLIMP5-PRINTS PROFILES SCAN HMMER-PFAM MOTIFS MOTIFS BLAST-DOMO BLIMP5-PRINTS PROFILES SCAN HMMER-PFAM MOTIFS SPSCAN
3	7474590CD1	497	S17 S286 S291 S3 S314 S356 S372 S375 S381 S382 S409 S440 S447 S5 S70 T225 T265 T427 T445 T461	N243		
4	7474594CD1	741	S397 S402 S471 S592 S641 S652 S656 S737 T237 T274 T292 T308 T388 T587	N119 N291		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7477585CD1	645	S251 S273 S277 S372 S414 S454 S47 S490 S522 S600 S64 S84 S97 T211 T302 T329 T340 T538 T547 Y368	N598 N71	PROTEIN KINASE DOMAIN DM00004   P51957   8-251: L35-S277 Tyrosine kinase catalytic domain PR00109: T108-Q121, Y148-L166, Y256-A278 Protein kinases signatures and profile protein_kinase_tyr.prf: Q134-S185 Eukaryotic protein kinase domain pkinase: Y29-L287 Protein_Kinase_St: I154-L166 PROTEIN KINASE DOMAIN DM00004   P53350   55-295: R99-A267, S253-L310 Tyrosine kinase catalytic domain PR00109: Y212-L230 Protein kinases signatures and profile protein_kinase_tyr.prf: E198-G250 Transmembrane domain transmem_domain: L555-S575 Eukaryotic protein kinase domain pkinase: Y97-A267, S268-F319 Protein_Kinase_Atp: I103-K126 Protein_Kinase_St: I218-L230	BLAST-DOMO BLIMPS-PRINTS PROFILES-SCAN HMMER-PFAM MOTIFS BLAST-DOMO BLIMPS-PRINTS PROFILES-SCAN HMMER HMMER-PFAM MOTIFS MOTIFS
6	7477587CD1	623	S32 S393 S439 S54 S61 S67 S80 S93 T195 T367 T454 T463 T584		PHORBOL-ESTER AND DAG BINDING DOMAIN DM01331   P49621   326-792: V321-K789 KINASE DIACYLGLYCEROL PHORBOLESTER BINDING TRANSFERASE DIGLYCERIDE DAG MULTIGENE FAMILY DGK PD002939: I575-P755 PROBABLE DIACYLGLYCEROL KINASE EC 2.7.1.107 DIGLYCERIDE DGK DAG HYPOTHETICAL PROTEIN TRANSFERASE CALCIUMBINDING PHORBOLESTER BINDING PD078865: A118-G236, L10-D85, T50-S81 KINASE DIACYLGLYCEROL PHORBOLESTER BINDING PROTEIN TRANSFERASE DIGLYCERIDE DAG MULTIGENE FAMILY PD002780: P431-W555	BLAST-DOMO BLAST-PRODROM BLAST-PRODROM BLAST-PRODROM BLAST-PRODROM
7	7594537CD1	797	S11 S136 S165 S208 S25 S294 S380 S670 S675 S684 S81 T2 T26 T274 T298 T312 T320 T388 T518 T62 T625 T689 T743 T762 T766 Y449	N546 N646 N793		BLAST-DOMO BLAST-PRODROM BLAST-PRODROM BLAST-PRODROM BLAST-PRODROM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DIACYLGLYCEROL KINASE, BETA EC 2.7.1.107 DIGLYCERIDE KINASE DGK 90 KD TRANSFERASE CALCIUMBINDING PHORBOLESTER BINDING MULTIGENE FAMILY PD119174: D352-H430 Diacylglycerol kinase catalytic domain PF00781: H331-Q336 P431-Y462 R483-L497 P509-Y532 K539-V559 N577-Y613 L655-G668 L747-Q758 Diacylglycerol kinase catalytic domain DAGKc: P431-W555 Diacylglycerol kinase accessory domain DAGKa: I575-P755 Phorbol esters/diacylglycerol binding domain d DAG_PE-gind: H238-C287, H303-C351 EF hand ehand: K146-M174, I191-T219 Phorbol esters/diacylglycerol binding domain BL00479: Q264-C279, L514-L526, H238-G260 Phorbol esters/diacylglycerol binding domain dag_pe_binding_domain.prf: Y250-G378 Dag_Pe_Binding_Domain: H238-C287 Ef_Hand: D155-L167, D200-W212	BLAST-PRODOM    BLIMPS-PFAM   HMMEP-PFAM HMMEP-PFAM HMMEP-PFAM BLIMPS-BLOCKS PROFILES CAN MOTIFS MOTIFS
8	70467491CD1	749	S141 S2 S24 S346 S374 S417 S424 S444 S456 S457 S461 S49 S494 S495 S516 S634 S653 S659 S664 S730 T118 T283 T302 T33 T36 T508 T512 T519 T535 T614 T618 T623 T82 T9 Y113	N386 N399 N400 N479 N533 N637	PROTEIN KINASE DOMAIN DM00004   P27448   58-297: L62-L303 KINASE SERINE/THREONINEPROTEIN PROTEIN TRANSFERASE ATPBINDING SERINE/THREONINE PUTATIVE KIN1 EMK PAR1 PD004300: G633-L749 KINASE SERINE/THREONINEPROTEIN SERINE/THREONINE PUTATIVE TRANSFERASE ATPBINDING PROTEIN EMK P78 CDC25C PD008571: S413-E632 KINASE SERINE/THREONINEPROTEIN PUTATIVE SERINE/THREONINE TRANSFERASE ATPBINDING PROTEIN PAR1 KP78 EMK PD005838: I312-R412 SERINE/THREONINE KINASE PD119193: S551-P622	BLAST-DOMO BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Tyrosine kinase catalytic domain PR00109: Y173-L191, V239-Q261 Protein kinases signatures and profile protein_kinase_tyr.prf: K122-G212 Eukaryotic protein kinase domain pkinase: Y60-E85 Eukaryotic protein kinase domain pkinase: F137-I312 Protein_Kinase_St: I179-L191 do CHOLINE; KINASE; YDR147W; B0285.10; DM01931 P35790 128-455: D258-K376, F131-P300 do CHOLINE; KINASE; YDR147W; B0285.10; DM01931 P46560 1-305: E125-A289 KINASE CHOLINE TRANSFERASE PROTEIN MULTIGENE FAMILY PUTATIVE LIKE CHROMOSOME III PD003547: V222-L382, V109-E240 KINASE TRANSFERASE CHOLINE PD02952: V243-I256, H263-N292 Choline/ethanolamine kinase Choline_kinase: T85-T356	BLIMPS-PRINTS PROFILES-SCAN HMMER-PFAM HMMER-PFAM MOTIFS BLAST-DOMO BLAST-DOMO BLAST-PRODOM BLIMPS-PRODOM HMMER-PFAM
9	7478559CD1	386	S237 S259 S355 S38 S380 T20 T322 T85 T93 Y271	N188		
10	1698381CD1	342	S180 S205 S238 S284 S288 S38 T247 Y15 Y211	N23	Eukaryotic protein kinase domain pkinase:Y4-F286, Protein kinases signatures and profile protein_kinase_tyrosine: E90-G154 PROTEIN KINASE DOMAIN DM00004 Q00532 7-278: K6-C277 PROTEIN KINASE DOMAIN DM00004 Q00526 6-286: K6-F286 PROTEIN KINASE DOMAIN DM00004 P23437 6-286: K6-G218 PROTEIN KINASE DOMAIN DM00004 P51958 6-277: K6-G218 KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: V161-F286	HMMER_PFAM PROFILES-SCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Tyrosine kinase catalytic domain signature PR00109: F116-I134	BLIMPS_PRINTS
11	7474637CD1	1164	S114 S119 S152 S258 S39 S399 S41 S432 S450 S511 S56 S586 S587 S591 S608 S623 S654 S66 S664 S695 S766 S820 S873 S958 S967 S1075 T316 T419 T486 T514 T518 T659 T678 T863 T908 T955 T1046 T1118	N124 N314 N651 N1059 N1122	Serine/Threonine protein kinases active-site signature C122-I134 Phorbol esters/diacylglycerol binding domain: F188-A259 signal_cleavage:M1-A32 Phorbol esters/diacylglycerol binding domain (C1 domain): H176-C225, H248-C298 PH domain: S66-T158 DAG kinase catalytic domain: P332-W457 DAG kinase accessory domain: V770-A927 PHORBOL-ESTER AND DAG BINDING DOMAIN DM01331 P49621 326-792: P332-H505, V770-E865, F869-L946, C279-L313, G198-C225 PHORBOL-ESTER AND DAG BINDING DOMAIN DM01331 Q09103 683-1148: V330-I459, T752-R948, C279-P310, A2-I61 PHORBOL-ESTER AND DAG BINDING DOMAIN DM01331 P23743 308-734: P332-L500, V770-F869, P872-L946 PHORBOL-ESTER AND DAG BINDING DOMAIN DM01331 I59282 352-782: C279-H505, V770-L946 KINASE DIACYLGLYCEROL ETA DIGLYCERIDE DAG TRANSFERASE PHORBOLESTER BINDING REPEAT MULTIGENE PD040467: S458-C769 DIACYLGLYCEROL PHORBOLESTER BINDING KINASE ETA DIGLYCERIDE DAG TRANSFERASE REPEAT MULTIGENE PD038733: A927-V1130 KINASE DIACYLGLYCEROL PHORBOLESTER BINDING TRANSFERASE DIGLYCERIDE DAG MULTIGENE FAMILY DKG PD002939: V770-E926 KINASE DIACYLGLYCEROL PHORBOLESTER BINDING PROTEIN TRANSFERASE DIGLYCERIDE DAG MULTIGENE FAMILY PD002780: V330-W457	MOTIFS PROFILES SCAN SPSCAN HMIMER_PFAM HMIMER_PFAM HMIMER_PFAM BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO: ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE; THREONINE; ATP; SERINE; DM06305 P32361 972-1114: Q813-L946 PROTEIN KINASE/ENDORIBONULCEASE PUTATIVE SERINE/THREONINE PROTEIN KINASE C41C4.4 CHROMOSOME II PRECURSOR TRANSFERASE PD152704: T197-L422, L88-E190 SERINE/THREONINE PROTEIN KINASE PRECURSOR TRANSMEMBRANE SIGNAL TRANSFERASE ATP-BINDING PROTEIN IRE1 GLYCOPROTEIN PD032590: W821-Y949 Tyrosine kinase catalytic domain signature PR00109: H666-I684, G721-L731, V743-D765 Serine/Threonine protein kinases active-site signature: I672-I684 Phosphorylase kinase family signature PR01049: P812-R823	BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLIMPS_PRINTS MOTIFS BLIMPS_PRINTS
14	1851973CD1	329	S264 S270 S293 S31 S311 S320 S7	N73	Eukaryotic protein kinase domain pkinase: F35-V180 Protein kinases signatures and profile protein_kinase_tyrosine: M132-R184 PROTEIN KINASE DOMAIN DM00004 P43565 796-1240: I37-R184 PROTEIN KINASE DOMAIN DM00004 A56155 714-1002: V38-L177 PROTEIN KINASE DOMAIN DM00004 P38679 238-527: V38-S178 PROTEIN KINASE DOMAIN DM00004 P53894 353-658: V38-S178 Tyrosine kinase catalytic domain signature PR00109: M110-H123, Y146-I164 Serine/Threonine protein kinases active-site signature: I152-I164	HMMER_PFAM PROFILES SCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLIMPS_PRINTS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7474604CD1	945	S110 S157 S193 S246 S289 S290 S31 S329 S356 S359 S405 S411 S611 S623 S636 S645 S67 S934 T170 T2 T217 T322 T42 T47 T496 T712 T839	N140 N155 N382 N631 N756 N888	Eukaryotic protein kinase domain pkinase: L661-M920 Protein kinases signatures and profile protein_kinase_tyrosine: K757-L801 PROTEIN KINASE DOMAIN DM00004 P27966 85-332: I663-F916 PROTEIN KINASE DOMAIN DM00004 P15056 458-705: I663-F916 PROTEIN KINASE DOMAIN DM00004 P10398 312-559: I667-F916 PROTEIN KINASE DOMAIN DM00004 B26126 305-552: I667-F916 KINASE SUPPRESSOR OF RAS1 KSR1 HB PROTEIN PD103125: V390-P557, K501-L661 L222-P323 KINASE SUPPRESSOR OF RAS KSR PHORBOLLESTER BINDING RAS1 KSR1 HB PD017776: L21-E344 S485-T519 Tyrosine kinase catalytic domain signature PR00109: Y771-Y789, W867-I877, M894-F916 Serine/Threonine protein kinases active-site signature: I777-Y789 Eukaryotic protein kinase domain pkinase: V645-H897 Ephrin receptor ligand binding domain EPH_lbd: E35-C211 Protein kinases signatures and profile protein_kinase_tyrosine: Q746-A799 RECEPTOR TYROSINE KINASE CLASS V DM00501 S51741 33-382: V36-G394 RECEPTOR TYROSINE KINASE CLASS V DM00501 P54759 33-382: V36-G394 RECEPTOR TYROSINE KINASE CLASS V DM00501 I48611 34-382: I37-G394 RECEPTOR TYROSINE KINASE CLASS V DM00501 I48612 34-382: I37-G394 KINASE RECEPTOR PRECURSOR TYROSINE PROTEIN EPHRIN TRANSFERASE ATP-BINDING PHOSPHORYLATION TRANSMEMBRANE GLYCOPROTEIN PD001495: E35-C211	HMMER_PFAM PROFILESSCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLIMPS_PRINTS MOTIFS HMMER_PFAM HMMER_PFAM PROFILESSCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM
16	7474721CD1	1009	S184 S203 S244 S293 S325 S44 S473 S62 S625 S682 S686 S805 S825 S851 S980 T108 T121 T133 T162 T214 T224 T232 T32 T423 T488 T551 T616 T619 Y504 Y766 Y801	N311 N486		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	74777421CD1	1009			KINASE RECEPTOR PRECURSOR TYROSINE PROTEIN EPHRIN TRANSFERASE ATP-BINDING PHOSPHORYLATION TRANSMEMBRANE GLYCOPROTEIN PD149648: A213-A284 EPH FAMILY PROTEIN PD002683: P339-T451 KINASE RECEPTOR PRECURSOR TYROSINE PROTEIN EPHRIN TRANSFERASE ATP-BINDING PHOSPHORYLATION TRANSMEMBRANE SIGNAL PD001551: C285-R336 Receptor tyrosine kinase BL00239: E694-Q741, L747-R769, A772-S797, E798-Y847, G852-I896 Receptor tyrosine kinase BL00790: L751-A772, S805-W837, E838-G862, F863-K911, A955-R998, E35-N56, D65-P116, K172-A225, P252-Q276, C282-P329, R351-L377, C390-S433 signal peptide: M1-A33 transmembrane domain: V568-W589 signal_cleavage: M1-A33	BLAST_PRODUM BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_BLOCKS HMMER HMMER
17	7478815CD1	917	S243 S364 S379 S449 S503 S547 S551 S772 S787 S791 S810 S826 S896 T114 T161 T172 T275 T35 T508 T523 T569 T625 T722 T726 T811 T877 Y27 Y497	N122 N208 N655	Hexokinase hexokinase:E16-V463 Q464-L910 Hexokinases signature hexokinases: I577-R642, V130-R195 HEXOKINASES DM00597 P27881 465-915: Q466-A913, D17-Q464 HEXOKINASES DM00597 P52789 465-915: Q466-A913, D17-Q464 HEXOKINASES DM00597 S48809 465-915: Q466-A913, D17-Q464 HEXOKINASES DM00597 P27595 465-915: Q466-Q911, D17-Q466 HEXOKINASE TRANSFERASE KINASE GLYCOLYSIS ATP-BINDING TYPE ALLOSTERIC ENZYME HK DUPLICATION PD001109: Q466-D886, E699-A907, E16-D439, D251-R462 Hexokinases proteins. BL00378: V22-K49, V509-I545, V207-G250, M255-D266, Y724-G769, S892-V906 Hexokinase family signature PR00475: L529-I545, L597-F622, I650-Y666, V226-E240, Q291-M313, V818-I840, M890-V906 Hexokinases L597-F622	SPSCAN HMMER_PPFAM PROFILESSCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7477141CD1	2380	S143 S166 S241 S277 S278 S285 S299 S343 S480 S537 S553 S568 S602 S711 S736 S996 S1033 S1035 S1037 S1062 S1127 S1523 S1571 S1245 S1435 S1468 S1506 S1586 S1609 S1679 S1691 S1747 S1117 S1527 S1557 S1578 S1594 S1613 S1736 S1747 S1876 S1947 S2137 S2171 S2253 S2321 S2058 S2062 S2165 S2269 S680 S754 S986 T108 T153 T158 T170 T350 T408 T476 T498 T578 T614 T692 T803 T862 T957 T1068 T1082 T1311 T1493 T1802 T1981 T2080 T1301 T1856 T1901 T2069 T2101 T2144 T2348 T1608 T2343 Y632 Y772 Y822	N37 N1675, N1847 N1874, N2099 N2299	Eukaryotic protein kinase domain pkinase: Y714-F967, Y2079-L2331 PROTEIN KINASE DOMAIN DM00004 S07571 5152- 5396: D715-D952, E2083-L2322 PROTEIN KINASE DOMAIN DM00004 P533355 15- 257: Q718-D952, E2083-L2322 PROTEIN KINASE DOMAIN DM00004 JN0583 727- 969: I716-D952, L2082-L2312 PROTEIN KINASE DOMAIN DM00004 P07313 298- 541: Q718-R953, G2088-S2321 Tyrosine kinase catalytic domain signature PR00109: Y822-V840 signal peptide: M52-A70 Eukaryotic protein kinase domain pkinase: Y2079-L2331 Protein kinases ATP-binding region signature: I720-K743 Serine/Threonine protein kinases active- site signature: V828-V840, V2194-L2206	HMMER_PFAM BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLIMPS_PRINTS HMMER HMMER_PFAM MOTIFS MOTIFS

19	2190612CD1	505	S100 S117 S160 S330 S419 S425 S437 S458 S69 S74 S82 T108 T26 T430 T58	N147	Eukaryotic protein kinase domain: Y128-V409 Protein kinases signatures and profile protein kinase tyrosine: Q251-N303 PROTEIN KINASE DOMAIN DM00004 A57156 130-399: L130-V400 PROTEIN KINASE DOMAIN DM00004 P50526 136-399: E133-I398 PROTEIN KINASE DOMAIN DM00004 P38990 135-438: E133-E320, N303-V400 PROTEIN KINASE DOMAIN DM00004 P43637 52-334: I134-I378	HMMER_PFAM PROFILES SCAN BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO
----	------------	-----	---	------	---	---

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7477549CD1	1572	S161 S280 S307 S363 S407 S430 S471 S545 S625 S629 S646 S675 S711 S730 S737 S807 S811 S815 S841 S1058 S1294 S1162 S1500 S1405 S1414 S1556 T455 T590 T673 T888 T956 T1088 T1378		KINASE PROTEIN BETA CA2+/CALMODULIN DEPENDENT CA+/CALMODULIN DEPENDENT CAM KINASE IV ISOFORM PHOSPHORYLASE B PD031900: M1-Q127	BLAST_PRODUM
					KINASE PROTEIN BETA CA2+/CALMODULIN DEPENDENT CA+/CALMODULIN DEPENDENT CAM KINASE IV ISOFORM PHOSPHORYLASE B PD019141: V409-F463	BLAST_PRODUM
					KINASE PROTEIN CA2+/CALMODULIN DEPENDENT IV ISOFORM PHOSPHORYLASE B GLYCOGEN SYNTHASE A PD027014: E464-S505	BLAST_PRODUM
					Tyrosine kinase catalytic domain signature PR00109: Y265-L283, G312-I322	BLIMPS_PRINTS
					ATP/GTP-binding site motif A (P-loop) G485-S492	MOTIFS
					Protein kinases ATP-binding region signature: I134-K157	MOTIFS
					Serine/Threonine protein kinases active-site signature: I271-L283	MOTIFS
					Phorbol esters diacylglycerol binding domain: C900-S963	PROFILES SCAN
					Eukaryotic protein kinase domain pkinase: F71-F337	HMMER_PPFAM
					PROTEIN KINASE DOMAIN DM00004 Q09013 83-336: I73-R325	BLAST_DOMO
					PROTEIN KINASE DOMAIN DM00004 S42867 75-498: I73-H252, V232-Y398	BLAST_DOMO
					PROTEIN KINASE DOMAIN DM00004 I38133 90-369: E72-L220, V232-G324	BLAST_DOMO
					PROTEIN KINASE DOMAIN DM00004 P53894 353-658: L74-G215, V232-R325	BLAST_DOMO
					PHORBOL ESTER BINDING KINASE DYSTROPHY KINASE RELATED CDC42 BINDING SIMILAR SERINE/THREONINE PROTEIN GENGHIS KHAN PD150840: W1355-G1462	BLAST_PRODUM
					PHORBOL ESTER BINDING KINASE DYSTROPHY KINASE RELATED CDC42 BINDING SIMILAR SERINE/THREONINE PROTEIN GENGHIS KHAN PD151400: T1039-R1140	BLAST_PRODUM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE RHO ASSOCIATED COILED COIL PROTEIN FORMING PHORBOL ESTER BINDING DYSTROPHY KINASE RELATED CDC42 BINDING PD006715: T944-V1038, H433-L456	BLAST_PRODUM
					PHORBOL ESTER BINDING DYSTROPHY KINASE RELATED CDC42 BINDING KINASE GENGHIS KHAN MYOTONIC MYOTONIC PD011252:S694-S815	BLAST_PRODUM
					Tyrosine kinase catalytic domain signature PR00109: C257-E279, M148-S161, S185-L203	BLIMPS_PRINTS
					Phorbol esters/diacylglycerol binding dom	HMMER_PFAM
					DAG_PE-bind: H887-C935	
					Phorbol esters/diacylglycerol binding domain: H887-C935,	MOTIFS
					Protein kinases ATP-binding region signature I77-K100	MOTIFS
					Serine/Threonine protein kinases active-site signature: Y191-L203	MOTIFS
					CNH domain: L1100-K1380	HMMER_PFAM
					Protein kinase C terminal domain: P351-D366	HMMER_PFAM
					PH domain PH: T956-R1074	HMMER_PFAM
					signal_cleavage: M1-S37	SPSCAN

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
21	2564295CB1	4298	701-1736, 1- 3536-3629, 1- 356, 2349- 2589, 3956- 4298, 2841- 3428	FL2564295_g7160581_000014_g 387060_1_15-16 FL2564295_g7160581_000014_g 387060_1_16-17 FL2564295_g7160581_000014_g 387060_1_7-8 55078393J1 FL2564295_g7160581_000014_g 387060_1_8-9 55078386J1 FL2564295_g7160581_000014_g 387060_1_18-19 2564295H1 (ADRETUT01) g186554_CD FL2564295_g7160581_000014_g 387060_1_9-10 387060_1_1-2 3599581H1 (DRGTNOT01) FL2564295_g7160581_000014_g 387060_1_1-2 FL2564295_g7160581_000014_g 387060_1_10-11 FL2564295_g7160581_000014_g 387060_1_20-21 FL2564295_g7160581_000014_g 387060_1_11-12 FL2564295_g7160581_000014_g 387060_1_2-3 FL2564295_g7160581_000014_g 387060_1_12-13 FL2564295_g7160581_000014_g 387060_1_3-4 FL2564295_g7160581_000014_g 387060_1_4-5	3200 3253 1938 37 2167 3594 1 4048 442 2335 3453 441 2531 3884 2573 994 2794 1298 1441	3482 3593 2334 717 2530 709 3883 4298 4250 2572 3756 1297 2793 4250 2930 1440 3093 1585 1800



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
21				FL2564295_g7160581_000014_g 387060_1_14-15 FL2564295_g7160581_000014_g 387060_1_19-20	3094 3754	3252 4018
22	2837050CB1	2863	1-430, 2346- 2863	6854541H1 (BRAIFEN08) g1164223 71191190V1 7728560H1 (UTRCDE01) 71972220V1 71972389V1 6881340H1 (BRAHTDR03) 7401101H1 (SINIDME01)	782 1 1439 79 2227 2180 1555 598	1467 496 2085 681 2863 2857 2209 1293
23	7474590CB1	1494	1-1494	GBI.g8103343_000001.edit FL7474590_g7630344_000002_g 6779549_1_1	1 1	1494 1116
24	7474594CB1	2341	682-792, 1- 262, 1522- 2341, 1254- 1373, 339-361	55053685J1 6949237H1 (BRAITDR02) 8016740J1 (BMAFXE01) GNN.g8247875_000031_002 7278940H1 (BMAFXE01) GNN.g6689704_000006_002	1512 858 340 1 1281 1180	2341 1544 959 426 1779 1590
25	7477585CB1	2552	1-465, 1075- 1150	71975408V1 55030002H1 55030074J1 1406660F6 (LATFTUT02) 6329987H1 (BRANDIN01) 71987367V1 6704049H1 (DRGCNOT02) 55030089H1 g8671962_edit 5823464F7 (PROSTUS23)	1988 612 1241 1 1384 2019 1849 679 1 1662	2534 1305 1900 686 1930 2552 2517 1390 1980 2164
26	7477587CB1	2176	1276-1873, 1- 286			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
27	7594537CB1	4277	2383-2614, 1- 611-1170, 1- 518, 2763- 2834, 1714- 1859, 3119- 4277	7594537H1 (LIVRNOC07) 7328693H1 (UTRCDIE01)	130 1	766 351
28	70467491CB1	2616	1717-2616, 1- 425	2395018F6 (THP1AZT01) FL70467491_g7708222_g759580 0	2015 1	2520 2250
29	7478559CB1	1253	1215-1253, 1- 53	g3770955 7661715J1 (OVARNOE02) g5769093	1 655 314	321 1253 804
30	1698381CB1	1790	1-146, 892- 1313, 1659- 1790, 186-237	1698381F6 (BLADTUT05) 55068293J1 71870273V1 1698381T6 (BLADTUT05)	523 1 1186 774	1019 786 1790 1363
31	7474637CB1	4132	3420-3535, 1- 377, 4035- 4132, 1301- 2486	4129796F6 (CARGDIT01) 55076747H1 55075847H1 55075848H1 55077477H1 GBI.g8247425_000008_000011. edit 55076756J1 GNN.g6648263_002.edit5p 5286993H2 (EPIPUNA01) 7721743H2 (THYRDIE01) 6766106H1 (BRAUNOR01) 55061367H1 6766106J1 (BRAUNOR01) 1752420H1 (LIVRTUT01) 55046242J2 3152909F6 (TLYMTXT02)	3639 2871 1379 1623 1045 504 3148 2805 1041 35 1893 2149 368 1 694 1	4132 3468 1783 1987 1472 1126 3745 3026 1168 503 2356 2841 909 157 1137 145
32	7170260CB1	1137	877-1137			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
32				7659273J1 (OVARNOE02) 55046250H1 3343082F7 (SPLNNOF09)	416 692 144	971 1108 555
33	1797506CB1	3365	1-1032, 3340- 3365, 1532- 1735	1513994T6 (PANCTUT01) FL11797506_g7458755_000012_g 3766209	2793 1	3365 2898
34	1851973CB1	2049	1-125, 1836- 2049, 806-915	7667239H1 (URETTUC01) 55075655J1 55077257J1 55067487H1 1454205F1 (PENITUT01) 1454205T6 (PENITUT01)	1289 547 378 1 1179 1436	1800 1222 1221 532 1617 2049
35	7474604CB1	2962	1-1526, 1757- 2114, 2481- 2962	55075789J1 8104459J1 (MIXDDIE02) 55056946J2 6884701F6 (BRAHTDR03) 55067076J1 55075383J1	1760 1 1734 2255 1214 651	2440 746 2433 2962 1763 1335
36	7474721CB1	3112	2395-3112, 1353-1459, 2014-2280	6802884F6 (COLENOR03) 71976507V1 55057353J1 GBI:g6996165_000001.raw GBI:g6996165.raw 55062828H1 71980671V1	2055 1564 314 1910 140 1 1418	2826 2315 980 3112 1735 712 2051
37	7478815CB1	3650	862-1366, 1826-1999, 1- 787, 3623- 3650	55076655H1 6934749H1 (SINTTMR02) 238539R6 (SINTNOT02) 614864T6 (COLNTUT02) 70845065V1 70842842V1	1 1710 3159 3004 1862 2420	658 2388 3647 3614 2441 3073

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
37				72026676V1 55075416J1 9657793 70863076V1 2605255F6 (LUNGUT07) 7355120H1 (HEARNON03) GBI:98014664 97242948_CD 3012344H1 (MUSCNOT07) 71179707V1 7642405J1 (SEMVTDE01)	966 358 919 2471 3378 7201 1 63 7488 6783 6728	1742 926 1011 3154 3650 7767 260 6763 7772 7436 7294
38	7477141CB1	7789	1-699, 6785- 6880, 7767- 7789, 7184- 7214, 1237- 6218	70775995V1 55024095J1 (PKINDNV04) 6854667H1 (BRAIFEN08) 7188730H2 (BRATD1C01) 70780513V1 70780809V1	1 914 1441 1353 500 384	498 1558 1937 1820 981 919
39	2190612CB1	1937	727-1188, 1- 643, 1731- 1761	55121415H1 55121423J1 7992167H1 (UTRSD1C01) 71999521V1 6822270H1 (SINTNOR01) GNN.g4755212_010.edit 6594083H1 (LUNGFER02) 7164493R8 (PLACNOR01) 71583419V1 7402224H1 (SINIDME01) 7694930H1 (LNODTUE01) 7978995H1 (LSUBDMC01)	4574 4413 3402 1448 857 1 2835 3204 713 289 1082 1478	5373 5274 4043 1590 1407 4567 3147 3711 1385 795 1448 2186
40	7477549CB1	5373	4983-5373, 1- 1612, 2046- 2470, 4414- 4442, 2596- 2647, 2814- 3056			

Table 5

Polynucleotide SEQ ID NO:	Incye Project ID	Representative Library
21	2564295CB1	ADRETUT01
22	2837050CB1	THYRN0T03
24	7474594CB1	BMARTXE01
25	7477585CB1	BRALNON02
26	7477587CB1	PROSTUS23
27	7594537CB1	LIVRNOC07
28	70467491CB1	PROSNOT18
29	7478559CB1	OVARNOE02
30	1698381CB1	BLADTUT05
31	7474637CB1	EPFIPUNA01
32	7170260CB1	OVARNOE02
33	1797506CB1	COLENOR03
34	1851973CB1	PENITUT01
35	7474604CB1	BRAHTDR03
36	7474721CB1	COLENOR03
37	7478815CB1	SINITUT03
38	7477141CB1	SKIRNOR01
39	2190612CB1	ADRETUT07
40	7477549CB1	SINTNOR01

Table 6

Library	Vector	Library Description
ADRETUT01	PSPORT	Library was constructed using RNA isolated from right adrenal tumor tissue removed from a 50-year-old Turkish male during a unilateral adrenalectomy. Pathology indicated a metastatic renal cell carcinoma that formed a circumscribed, spongy, hemorrhagic nodule situated in the region of the medulla. The patient presented with corticoadrenal insufficiency, incisional hernia, and non-alcoholic steato hepatitis. Patient history included renal cell carcinoma. Family history included liver cancer.
ADRETUT07	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 43-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated pheochromocytoma.
BLADTUT05	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microm for 8 hours.
BRAHTDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from archaecortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAIINON02	pINCY	This thalamus tissue library was constructed from 4.24 million independent clones from a thalamus tissue library. Starting RNA was made from thalamus tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. Scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding

Table 6 (cont.)

Library	Vector	Library Description
		gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
COLENOE03	PCDNA2.1	Library was constructed using RNA isolated from colon epithelium tissue removed from a 13-year-old Caucasian female who died from a motor vehicle accident.
EPIFUNA01	PSPORT	Library was constructed using RNA isolated from untreated prostatic epithelial cell tissue removed from a 17-year-old Hispanic male. Serologies were negative.
LIVNOC07	pINCY	Library was constructed using pooled cDNA from two different donors. cDNA was generated using RNA isolated from liver tissue removed from a 20-week-old Caucasian male fetus who died from Patau's Syndrome (donor A) and a 16-week-old Caucasian female fetus who died from anencephaly (donor B). Family history included mitral valve prolapse in the mother of donor B.
OVARNOE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from right ovary tissue removed from a 47-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, incisional hernia repair, and panniculectomy. The patient presented with premenopausal menorrhagia. Patient history included osteoarthritis, tubal pregnancy, and polio osteopathy of the left leg. Previous surgeries included gastroenterostomy, plastic repair of the palate, adenotonsillectomy, dilation and curettage, cholecystectomy, and bladder reconstruction. Patient medications included vitamins, iron, and zinc. Family history included benign hypertension and type II diabetes in the father; and type II diabetes in the sibling(s).
PENITUT01	pINCY	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.

Table 6 (cont.)

Library	Vector	Library Description
PROSNOT18	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrectomy. Pathology indicated adenofibromatous hyperplasia; this tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
PROSTUS23	pINCY	This substracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.
SKIRNOR01	PCDNA2.1	Random-primed library was constructed using RNA isolated from skin tissue removed from the breast of a 17-year-old Caucasian female during bilateral reduction mammoplasty. Patient history included breast hypertrophy. Family history included benign hypertension.
SINITUT03	pINCY	Library was constructed using RNA isolated from ileal tumor tissue obtained from a 49-year-old Caucasian female during destruction of peritoneal tissue, peritoneal adhesiolysis, ileum resection, and permanent colostomy. Pathology indicated grade 4 adenocarcinoma. Patient history included benign hypertension. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, regional lymph node excision, an incidental appendectomy, and dilation and curettage. Family history included benign hypertension, cerebrovascular disease, hyperlipidemia, atherosclerotic coronary artery disease, hyperlipidemia, type II diabetes, and stomach cancer.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.



Table 6 ( cont.)

Library	Vector	Library Description
THYRNOT03	pTNCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta B value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx B value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Altwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribkov, M. et al. (1988) CABIOS 4:61-66; Gribkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-20,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and  
10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-20.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID  
20 NO:21-40.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell  
30 is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
  - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting
  - 5 of SEQ ID NO:21-40,
  - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,
  - c) a polynucleotide complementary to a polynucleotide of a),
  - d) a polynucleotide complementary to a polynucleotide of b), and
  - 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
  - 20 is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

5

18. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 16.

10 19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15 20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

20 21. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 25 b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30 24. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5

29. A diagnostic test for a condition or disease associated with the expression of PKIN in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

10 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- 15 b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

20 31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

25 33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

30 35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit



an antibody response;

- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-20 in a sample, comprising the steps of:

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
  - b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide
- 5 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 from a sample, the method comprising:

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding
- 10 of the antibody and the polypeptide; and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

15 45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

20 48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

25 50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

30 53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
- 5 57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
- 10 60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
- 15 62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
- 20 65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
- 25 67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.
- 30 70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.
71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.
73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.
- 5 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.
75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.
76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.
- 10 77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.
78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.
- 15 79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.
80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.
81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.
- 20 82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.
83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.
- 25 84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

<110> INCYTE GENOMICS, INC.

YUE, Henry

KHAN, Farrah A.

GURURAJAN, Rajagopal

HAFALIA, April J. A.

WALIA, Narinder K.

PATTERSON, Chandra

RAMKUMAR, Jayalaxmi

GANDHI, Ameena R.

POLICKY, Jennifer L.

BAUGHN, Mariah R.

TRIBOULEY, Catherine M.

THORNTON, Michael

BANDMAN, Olga

NGUYEN, Danniel B.

LU, Yan

BURFORD, Neil

LAL, Preeti

DING, Li

YAO, Monique G.

ELLIOTT, Vicki S.

RECIPON, Shirley A.

KEARNEY, Liam

LU, Dyung Aina M.

GREENWALD, Sara R.

TANG, Y. Tom

XU, Yuming

WALSH, Roderick T.

GIETZEN, Kimberly J.

YANG, Junming

HILLMAN, Jennifer L.

<120> HUMAN KINASES

<130> PI-0162 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/220,038; 60/222,112; 60/222,831; 60/224,729

<151> 2000-07-21; 2000-07-28; 2000-08-04; 2000-08-11

<160> 40

<170> PERL Program

<210> 1

<211> 1297

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2564295CD1

<400> 1

Met Ala Val Pro Ser Leu Trp Pro Trp Gly Ala Cys Leu Pro Val

	20		25		30
Pro Ser Leu Asp	Ile Arg Ser Glu Val	Ala Glu Leu Arg Gln Leu			
	35		40		45
Glu Asn Cys Ser	Val Val Glu Gly His	Leu Gln Ile Leu Leu Met			
	50		55		60
Phe Thr Ala Thr	Gly Glu Asp Phe Arg	Gly Leu Ser Phe Pro Arg			
	65		70		75
Leu Thr Gln Val	Thr Asp Tyr Leu Leu	Leu Phe Arg Val Tyr Gly			
	80		85		90
Leu Glu Ser Leu	Arg Asp Leu Phe Pro	Asn Leu Ala Val Ile Arg			
	95		100		105
Gly Thr Arg Leu	Phe Leu Gly Tyr Ala	Leu Val Ile Phe Glu Met			
	110		115		120
Pro His Leu Arg	Asp Val Ala Leu Pro	Ala Leu Gly Ala Val Leu			
	125		130		135
Arg Gly Ala Val	Arg Val Glu Lys Asn	Gln Glu Leu Cys His Leu			
	140		145		150
Ser Thr Ile Asp	Trp Gly Leu Leu Gln	Pro Ala Pro Gly Ala Asn			
	155		160		165
His Ile Val Gly	Asn Lys Leu Gly Glu	Glu Cys Ala Asp Val Cys			
	170		175		180
Pro Gly Val Leu	Gly Ala Ala Gly Glu	Pro Cys Ala Lys Thr Thr			
	185		190		195
Phe Ser Gly His	Thr Asp Tyr Arg Cys	Trp Thr Ser Ser His Cys			
	200		205		210
Gln Arg Val Cys	Pro Cys Pro His Gly	Met Ala Cys Thr Ala Arg			
	215		220		225
Gly Glu Cys Cys	His Thr Glu Cys Leu	Gly Gly Cys Ser Gln Pro			
	230		235		240
Glu Asp Pro Arg	Ala Cys Val Ala Cys	Arg His Leu Tyr Phe Gln			
	245		250		255
Gly Ala Cys Leu	Trp Ala Cys Pro Pro	Gly Thr Tyr Gln Tyr Glu			
	260		265		270
Ser Trp Arg Cys	Val Thr Ala Glu Arg	Cys Ala Ser Leu His Ser			
	275		280		285
Val Pro Gly Arg	Ala Ser Thr Phe Gly	Ile His Gln Gly Ser Cys			
	290		295		300
Leu Ala Gln Cys	Pro Ser Gly Phe Thr	Arg Asn Ser Ser Ser Ile			
	305		310		315
Phe Cys His Lys	Cys Glu Gly Leu Cys	Pro Lys Glu Cys Lys Val			
	320		325		330
Gly Thr Lys Thr	Ile Asp Ser Ile Gln	Ala Ala Gln Asp Leu Val			
	335		340		345
Gly Cys Thr His	Val Glu Gly Ser Leu	Ile Leu Asn Leu Arg Gln			
	350		355		360
Gly Tyr Asn Leu	Glu Pro Gln Leu Gln	His Ser Leu Gly Leu Val			
	365		370		375
Glu Thr Ile Thr	Gly Phe Leu Lys Ile	Lys His Ser Phe Ala Leu			
	380		385		390
Val Ser Leu Gly	Phe Phe Lys Asn Leu	Lys Leu Ile Arg Gly Asp			
	395		400		405
Ala Met Val Asp	Gly Asn Tyr Thr Leu	Tyr Val Leu Asp Asn Gln			
	410		415		420
Asn Leu Gln Gln	Leu Gly Ser Trp Val	Ala Ala Gly Leu Thr Ile			
	425		430		435
Pro Val Gly Lys	Ile Tyr Phe Ala Phe	Asn Pro Arg Leu Cys Leu			

	440		445		450
Glu His Ile Tyr Arg Leu Glu Glu Val Thr Gly Thr Arg Gly Arg					
	455		460		465
Gln Asn Lys Ala Glu Ile Asn Pro Arg Thr Asn Gly Asp Arg Ala					
	470		475		480
Ala Cys Gln Thr Arg Thr Leu Arg Phe Val Ser Asn Val Thr Glu					
	485		490		495
Ala Asp Arg Ile Leu Leu Arg Trp Glu Arg Tyr Glu Pro Leu Glu					
	500		505		510
Ala Arg Asp Leu Leu Ser Phe Ile Val Tyr Tyr Lys Glu Ser Pro					
	515		520		525
Phe Gln Asn Ala Thr Glu His Val Gly Pro Asp Ala Cys Gly Thr					
	530		535		540
Gln Ser Trp Asn Leu Leu Asp Val Glu Leu Pro Leu Ser Arg Thr					
	545		550		555
Gln Glu Pro Gly Val Thr Leu Ala Ser Leu Lys Pro Trp Thr Gln					
	560		565		570
Tyr Ala Val Phe Val Arg Ala Ile Thr Leu Thr Thr Glu Glu Asp					
	575		580		585
Ser Pro His Gln Gly Ala Gln Ser Pro Ile Val Tyr Leu Arg Thr					
	590		595		600
Leu Pro Ala Ala Pro Thr Val Pro Gln Asp Val Ile Ser Thr Ser					
	605		610		615
Asn Ser Ser Ser His Leu Leu Val Arg Trp Lys Pro Pro Thr Gln					
	620		625		630
Arg Asn Gly Asn Leu Thr Tyr Tyr Leu Val Leu Trp Gln Arg Leu					
	635		640		645
Ala Glu Asp Gly Asp Leu Tyr Leu Asn Asp Tyr Cys His Arg Gly					
	650		655		660
Leu Arg Leu Pro Thr Ser Asn Asn Asp Pro Arg Phe Asp Gly Glu					
	665		670		675
Asp Gly Asp Pro Glu Ala Glu Met Glu Ser Asp Cys Cys Pro Cys					
	680		685		690
Gln His Pro Pro Pro Gly Gln Val Leu Pro Pro Leu Glu Ala Gln					
	695		700		705
Glu Ala Ser Phe Gln Lys Lys Phe Glu Asn Phe Leu His Asn Ala					
	710		715		720
Ile Thr Ile Pro Ile Ser Pro Trp Lys Val Thr Ser Ile Asn Lys					
	725		730		735
Ser Pro Gln Arg Asp Ser Gly Arg His Arg Arg Ala Ala Gly Pro					
	740		745		750
Leu Arg Leu Gly Gly Asn Ser Ser Asp Phe Glu Ile Gln Glu Asp					
	755		760		765
Lys Val Pro Arg Glu Arg Ala Val Leu Ser Gly Leu Arg His Phe					
	770		775		780
Thr Glu Tyr Arg Ile Asp Ile His Ala Cys Asn His Ala Ala His					
	785		790		795
Thr Val Gly Cys Ser Ala Ala Thr Phe Val Phe Ala Arg Thr Met					
	800		805		810
Pro His Arg Glu Ala Asp Gly Ile Pro Gly Lys Val Ala Trp Glu					
	815		820		825
Ala Ser Ser Lys Asn Ser Val Leu Leu Arg Trp Leu Glu Pro Pro					
	830		835		840
Asp Pro Asn Gly Leu Ile Leu Lys Tyr Glu Ile Lys Tyr Arg Arg					
	845		850		855
Leu Gly Glu Glu Ala Thr Val Leu Cys Val Ser Arg Leu Arg Tyr					

860	865	870
Ala Lys Phe Gly Gly Val His Leu	Ala Leu Leu Pro Pro Gly Asn	
875	880	885
Tyr Ser Ala Arg Val Arg Ala Thr Ser	Leu Ala Gly Asn Gly Ser	
890	895	900
Trp Thr Asp Ser Val Ala Phe Tyr Ile	Leu Gly Pro Glu Glu Glu	
905	910	915
Asp Ala Gly Gly Leu His Val Leu Leu Thr	Ala Thr Pro Val Gly	
920	925	930
Leu Thr Leu Leu Ile Val Leu Ala Ala Leu	Gly Phe Phe Tyr Gly	
935	940	945
Lys Lys Arg Asn Arg Thr Leu Tyr Ala Ser	Val Asn Pro Glu Tyr	
950	955	960
Phe Ser Ala Ser Asp Met Tyr Val Pro Asp	Glu Trp Glu Val Pro	
965	970	975
Arg Glu Gln Ile Ser Ile Ile Arg Glu Leu	Gly Gln Gly Ser Phe	
980	985	990
Gly Met Val Tyr Glu Gly Leu Ala Arg Gly	Leu Glu Ala Gly Glu	
995	1000	1005
Glu Ser Thr Pro Val Ala Leu Lys Thr Val	Asn Glu Leu Ala Ser	
1010	1015	1020
Pro Arg Glu Cys Ile Glu Phe Leu Lys Glu	Ala Ser Val Met Lys	
1025	1030	1035
Ala Phe Lys Cys His Val Val Arg Leu Leu	Gly Val Val Ser	
1040	1045	1050
Gln Gly Gln Pro Thr Leu Val Ile Met Glu	Leu Met Thr Arg Gly	
1055	1060	1065
Asp Leu Lys Ser His Leu Arg Ser Leu Arg	Pro Glu Ala Glu Asn	
1070	1075	1080
Asn Pro Gly Leu Pro Gln Pro Ala Leu Gly	Glu Met Ile Gln Met	
1085	1090	1095
Ala Gly Glu Ile Ala Asp Gly Met Ala Tyr	Leu Ala Ala Asn Lys	
1100	1105	1110
Phe Val His Arg Asp Leu Ala Ala Arg Asn	Cys Met Val Ser Gln	
1115	1120	1125
Asp Phe Thr Val Lys Ile Gly Asp Phe Gly	Met Thr Arg Asp Val	
1130	1135	1140
Tyr Glu Thr Asp Tyr Tyr Arg Lys Gly Gly	Lys Gly Leu Leu Pro	
1145	1150	1155
Val Arg Trp Met Ala Pro Glu Ser Leu Lys	Asp Gly Ile Phe Thr	
1160	1165	1170
Thr His Ser Asp Val Trp Ser Phe Gly Val	Val Leu Trp Glu Ile	
1175	1180	1185
Val Thr Leu Ala Glu Gln Pro Tyr Gln Gly	Leu Ser Asn Glu Gln	
1190	1195	1200
Val Leu Lys Phe Val Met Asp Gly Gly Val	Leu Glu Glu Leu Glu	
1205	1210	1215
Gly Cys Pro Leu Gln Leu Gln Glu Leu Met	Ser Arg Cys Trp Gln	
1220	1225	1230
Pro Asn Pro Arg Leu Arg Pro Ser Phe Thr	His Ile Leu Asp Ser	
1235	1240	1245
Ile Gln Glu Glu Leu Arg Pro Ser Phe Arg	Leu Leu Ser Phe Tyr	
1250	1255	1260
Tyr Ser Pro Glu Cys Arg Gly Ala Arg Gly	Ser Leu Pro Thr Thr	
1265	1270	1275
Asp Ala Glu Pro Asp Ser Ser Pro Thr Pro	Arg Asp Cys Ser Pro	



1280 1285 1290  
 Gln Asn Gly Gly Pro Gly His  
 1295

<210> 2  
 <211> 718  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2837050CD1

<400> 2  
 Met Met Glu Glu Leu His Ser Leu Asp Pro Arg Arg Gln Glu Leu  
 1 5 10 15  
 Leu Glu Ala Arg Phe Thr Arg Val Gly Val Ser Lys Gly Pro Leu  
 20 25 30  
 Asn Ser Glu Ser Ser Asn Gln Ser Leu Cys Ser Val Gly Ser Leu  
 35 40 45  
 Ser Asp Lys Glu Val Glu Thr Pro Glu Lys Lys Gln Asn Asp Gln  
 50 55 60  
 Arg Asn Arg Lys Arg Lys Ala Glu Pro Tyr Glu Thr Ser Gln Gly  
 65 70 75  
 Lys Gly Thr Pro Arg Gly His Lys Ile Ser Asp Tyr Phe Glu Arg  
 80 85 90  
 Arg Val Glu Gln Pro Leu Tyr Gly Leu Asp Gly Ser Ala Ala Lys  
 95 100 105  
 Glu Ala Thr Glu Glu Gln Ser Ala Leu Pro Thr Leu Met Ser Val  
 110 115 120  
 Met Leu Ala Lys Pro Arg Leu Asp Thr Glu His Val Ala Gln Arg  
 125 130 135  
 Gly Ala Gly Leu Cys Phe Thr Phe Val Ser Ala Gln Gln Asn Ser  
 140 145 150  
 Pro Ser Ser Thr Gly Ser Gly Asn Thr Glu His Ser Cys Ser Ser  
 155 160 165  
 Gln Lys Gln Ile Ser Ile Gln His Arg Gln Thr Gln Ser Asp Leu  
 170 175 180  
 Thr Ile Glu Lys Ile Ser Ala Leu Glu Asn Ser Lys Asn Ser Asp  
 185 190 195  
 Leu Glu Lys Lys Glu Gly Arg Ile Asp Asp Leu Leu Arg Ala Asn  
 200 205 210  
 Cys Asp Leu Arg Arg Gln Ile Asp Glu Gln Gln Lys Met Leu Glu  
 215 220 225  
 Lys Tyr Lys Glu Arg Leu Asn Arg Cys Val Thr Met Ser Lys Lys  
 230 235 240  
 Leu Leu Ile Glu Lys Ser Lys Gln Glu Lys Met Ala Cys Arg Asp  
 245 250 255  
 Lys Ser Met Gln Asp Arg Leu Arg Leu Gly His Phe Thr Thr Val  
 260 265 270  
 Arg His Gly Ala Ser Phe Thr Glu Gln Trp Thr Asp Gly Tyr Ala  
 275 280 285  
 Phe Gln Asn Leu Ile Lys Gln Gln Glu Arg Ile Asn Ser Gln Arg  
 290 295 300  
 Glu Glu Ile Glu Arg Gln Arg Lys Met Leu Ala Lys Arg Lys Pro  
 305 310 315

<210> 3

&lt;211&gt; 497

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474590CD1

&lt;400&gt; 3

```

Met Tyr Ser Asp Ser Glu Asp Glu Ser Ser Glu Leu Ser Thr Val
 1          5          10          15
Leu Ser Met Phe Glu Glu Lys Glu Phe Thr Arg Gln Tyr Thr Val
          20          25          30
Leu Lys Thr Leu Ser Gln His Gly Thr Thr Glu Val Arg Leu Cys
          35          40          45
Ser His His Leu Thr Gly Val Thr Val Ala Val Lys Ala Leu Lys
          50          55          60
Tyr Gln Arg Trp Trp Glu Pro Lys Val Ser Glu Val Glu Ile Met
          65          70          75
Lys Met Leu Ser His Pro Asn Ile Val Ser Leu Leu Gln Val Ile
          80          85          90
Glu Thr Glu Gln Asn Ile Tyr Leu Ile Met Glu Val Ala Gln Gly
          95          100          105
Thr Gln Leu His Asn Arg Val Gln Glu Ala Arg Cys Leu Lys Glu
          110          115          120
Asp Glu Ala Arg Ser Ile Phe Val Gln Leu Leu Ser Ala Ile Gly
          125          130          135
Tyr Cys His Gly Glu Gly Val Val His Arg Asp Leu Lys Pro Asp
          140          145          150
Asn Val Ile Val Asp Glu His Gly Asn Val Lys Ile Val Asp Phe
          155          160          165
Gly Leu Gly Ala Arg Phe Met Pro Gly Gln Lys Leu Glu Arg Leu
          170          175          180
Cys Gly Ala Phe Gln Phe Ile Pro Pro Glu Ile Phe Leu Gly Leu
          185          190          195
Pro Tyr Asp Gly Pro Lys Val Asp Ile Trp Ala Leu Gly Val Leu
          200          205          210
Leu Tyr Tyr Met Val Thr Gly Ile Phe Pro Phe Val Gly Ser Thr
          215          220          225
Leu Ser Glu Ile Ser Lys Glu Val Leu Gln Gly Arg Tyr Glu Ile
          230          235          240
Pro Tyr Asn Leu Ser Lys Asp Leu Arg Ser Met Ile Gly Leu Leu
          245          250          255
Leu Ala Thr Asn Ala Arg Gln Arg Pro Thr Ala Gln Asp Leu Leu
          260          265          270
Ser His Pro Trp Leu Gln Glu Gly Glu Lys Thr Ile Thr Phe His
          275          280          285
Ser Asn Gly Asp Thr Ser Phe Pro Asp Pro Asp Ile Met Ala Ala
          290          295          300
Met Lys Asn Ile Gly Phe His Val Gln Asp Ile Arg Glu Ser Leu
          305          310          315
Lys His Arg Lys Phe Asp Glu Thr Met Ala Thr Tyr Asn Leu Leu
          320          325          330
Arg Ala Glu Ala Cys Gln Asp Asp Gly Asn Tyr Val Gln Thr Lys
          335          340          345
Leu Met Asn Pro Gly Met Pro Pro Phe Pro Ser Val Thr Asp Ser

```

	350		355		360
Gly Ala Phe Ser	Leu Pro Pro Arg Arg	Arg Ala Ser Glu Pro	Ser		
	365		370		375
Phe Lys Val Leu Val	Ser Ser Thr Glu Glu	His Gln Leu Arg	Gln		
	380		385		390
Thr Gly Gly Thr	Asn Ala Pro Phe Pro	Pro Lys Lys Thr Pro	Thr		
	395		400		405
Met Gly Arg Ser	Gln Lys Gln Lys Arg	Ala Met Thr Ala Pro	Cys		
	410		415		420
Ile Cys Leu Leu	Arg Asn Thr Tyr Ile	Asp Thr Glu Asp Ser	Ser		
	425		430		435
Phe Cys Thr Ser	Ser Gln Ala Glu Lys	Thr Ser Ser Asp Pro	Glu		
	440		445		450
Lys Ser Glu Thr	Ser Thr Ser Cys Pro	Leu Thr Pro Arg Gly	Trp		
	455		460		465
Arg Lys Trp Lys	Lys Arg Ile Val Ala	Cys Ile Gln Thr Leu	Cys		
	470		475		480
Cys Cys Thr Leu	Pro Gln Lys Lys Cys	Pro Arg Ser Val His	Pro		
	485		490		495
Gln Lys					

&lt;210&gt; 4

&lt;211&gt; 741

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; InCyte ID No: 7474594CD1

&lt;400&gt; 4

Met Ser Gly Leu Val	Leu Met Leu Ala Ala	Arg Cys Ile Val Gly
1	5	10 15
Ser Ser Pro Leu Cys	Arg Cys Arg Arg Arg	Pro Arg Arg Ile
	20	25 30
Gly Ala Gly Pro Gly	Arg Asp Asp Pro Gly	Arg Lys Ala Ala Ala
	35	40 45
Ala Gly Gly Ser Gly	Ser Pro Asn Ala Ala	Leu Ser Arg Pro Arg
	50	55 60
Pro Ala Pro Ala Pro	Gly Asp Ala Pro Pro	Arg Ala Ala Ala Ser
	65	70 75
Ala Ala Ala Ala Ala	Ala Ala Ala Ala Gly	Thr Glu Gln Val Asp
	80	85 90
Gly Pro Leu Arg Ala	Gly Pro Ala Asp Thr	Pro Pro Ser Gly Trp
	95	100 105
Arg Met Gln Cys Leu	Ala Ala Ala Leu Lys	Asp Glu Thr Asn Met
	110	115 120
Ser Gly Gly Gly Glu	Gln Ala Asp Ile Leu	Pro Ala Asn Tyr Val
	125	130 135
Val Lys Asp Arg Trp	Lys Val Leu Lys Lys	Ile Gly Gly Gly Gly
	140	145 150
Phe Gly Glu Ile Tyr	Glu Ala Met Asp Leu	Leu Thr Arg Glu Asn
	155	160 165
Val Ala Leu Lys Val	Glu Ser Ala Gln Gln	Pro Lys Gln Val Leu
	170	175 180

Lys Met Glu Val	Ala Val Leu Lys Lys	Leu Gln Gly Lys Asp His	185	190	195
Val Cys Arg Phe	Ile Gly Cys Gly Arg	Asn Glu Lys Phe Asn Tyr	200	205	210
Val Val Met Gln	Leu Gln Gly Arg Asn	Leu Ala Asp Leu Arg Arg	215	220	225
Ser Gln Pro Arg	Gly Thr Phe Thr Leu	Ser Thr Thr Leu Arg Leu	230	235	240
Gly Lys Gln Ile	Leu Glu Ser Ile Glu	Ala Ile His Ser Val Gly	245	250	255
Phe Leu His Arg	Asp Ile Lys Pro Ser	Asn Phe Ala Met Gly Arg	260	265	270
Leu Pro Ser Thr	Tyr Arg Lys Cys Tyr	Met Leu Asp Phe Gly Leu	275	280	285
Ala Arg Gln Tyr	Thr Asn Thr Thr Gly	Asp Val Arg Pro Pro Arg	290	295	300
Asn Val Ala Gly	Phe Arg Gly Thr Val	Arg Tyr Ala Ser Val Asn	305	310	315
Ala His Lys Asn	Arg Glu Met Gly Arg	His Asp Asp Leu Trp Ser	320	325	330
Leu Phe Tyr Met	Leu Val Glu Phe Ala	Val Gly Gln Leu Pro Trp	335	340	345
Arg Lys Ile Lys	Asp Lys Glu Gln Val	Gly Met Ile Lys Glu Lys	350	355	360
Tyr Glu His Arg	Met Leu Leu Lys His	Met Pro Ser Glu Phe His	365	370	375
Leu Phe Leu Asp	His Ile Ala Ser Leu	Asp Tyr Phe Thr Lys Pro	380	385	390
Asp Tyr Gln Leu	Ile Met Ser Val Phe	Glu Asn Ser Met Lys Glu	395	400	405
Arg Gly Ile Ala	Glu Asn Glu Ala Phe	Asp Trp Glu Lys Ala Gly	410	415	420
Thr Asp Ala Leu	Leu Ser Thr Ser Thr	Ser Thr Pro Pro Gln Gln	425	430	435
Asn Thr Arg Gln	Thr Ala Ala Met Phe	Gly Val Val Asn Val Thr	440	445	450
Pro Val Pro Gly	Asp Leu Leu Arg Glu	Asn Thr Glu Asp Val Leu	455	460	465
Gln Gly Glu His	Leu Ser Asp Gln Glu	Asn Ala Pro Pro Ile Leu	470	475	480
Pro Gly Arg Pro	Ser Glu Gly Leu Gly	Pro Ser Pro His Leu Val	485	490	495
Pro His Pro Gly	Gly Pro Glu Ala Glu	Val Trp Glu Glu Thr Asp	500	505	510
Val Asn Arg Asn	Lys Leu Arg Ile Asn	Ile Gly Lys Val Thr Ala	515	520	525
Ala Arg Ala Lys	Gly Val Gly Gly Leu	Phe Ser His Pro Arg Phe	530	535	540
Pro Ala Leu Cys	Pro Cys Pro Val Pro	Pro Lys His Pro Val Pro	545	550	555
Gly His Leu Pro	Ala Cys Pro Ala Ser	Val Ser Arg Ser Leu Pro	560	565	570
Ala Leu Ala Ser	Leu Cys Leu Pro Ser	Ser Ser Ser Ser Val Ser	575	580	585
Phe Thr Leu Arg	Pro Ser Ala His	Ser Arg Leu Ile Ser Pro	590	595	600

Ser	Ser	Trp	His	Ser	Pro	Leu	Leu	Gln	Ser	Pro	Cys	Val	Glu	Glu
				605					610					615
Glu	Gln	Ser	Arg	Gly	Met	Gly	Val	Pro	Ser	Ser	Pro	Val	Arg	Ala
				620					625					630
Pro	Pro	Asp	Ser	Pro	Thr	Thr	Pro	Val	Arg	Ser	Leu	Arg	Tyr	Arg
				635					640					645
Arg	Val	Asn	Ser	Pro	Glu	Ser	Glu	Arg	Leu	Ser	Thr	Ala	Asp	Gly
				650					655					660
Arg	Val	Glu	Leu	Pro	Glu	Arg	Arg	Trp	Val	Trp	Gly	Gln	Gly	His
				665					670					675
Gly	Trp	Gly	Pro	Arg	Pro	Ser	Pro	Pro	Ser	Arg	Gly	Trp	Ser	Gly
				680					685					690
Gly	Lys	Val	Arg	Cys	Val	Ala	Glu	Val	Gly	Arg	Pro	Trp	Glu	Val
				695					700					705
Leu	Arg	Gly	Leu	Tyr	Leu	Gly	Leu	Gly	Ser	Asp	Ser	Val	Gly	Ala
				710					715					720
Arg	Asp	Arg	Ala	Trp	Glu	Asn	Gln	Trp	Gly	Ile	Gln	Arg	Gly	Pro
				725					730					735
Gly	Ser	Cys	Gln	Glu	Thr									
				740										

&lt;210&gt; 5

&lt;211&gt; 645

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477585CD1

&lt;400&gt; 5

Met	Leu	Lys	Phe	Gln	Glu	Ala	Ala	Lys	Cys	Val	Ser	Gly	Ser	Thr
1				5					10					15
Ala	Ile	Ser	Thr	Tyr	Pro	Lys	Thr	Leu	Ile	Ala	Arg	Arg	Tyr	Val
				20					25					30
Leu	Gln	Gln	Lys	Leu	Gly	Ser	Gly	Ser	Phe	Gly	Thr	Val	Tyr	Leu
				35					40					45
Val	Ser	Asp	Lys	Lys	Ala	Lys	Arg	Gly	Glu	Glu	Leu	Lys	Val	Leu
				50					55					60
Lys	Glu	Ile	Ser	Val	Gly	Glu	Leu	Asn	Pro	Asn	Glu	Thr	Val	Gln
				65					70					75
Ala	Asn	Leu	Glu	Ala	Gln	Leu	Leu	Ser	Lys	Leu	Asp	His	Pro	Ala
				80					85					90
Ile	Val	Lys	Phe	His	Ala	Ser	Phe	Val	Glu	Gln	Asp	Asn	Phe	Cys
				95					100					105
Ile	Ile	Thr	Glu	Tyr	Cys	Glu	Gly	Arg	Asp	Leu	Asp	Asp	Lys	Ile
				110					115					120
Gln	Glu	Tyr	Lys	Gln	Ala	Gly	Lys	Ile	Phe	Pro	Glu	Asn	Gln	Ile
				125					130					135
Ile	Glu	Trp	Phe	Ile	Gln	Leu	Leu	Leu	Gly	Val	Asp	Tyr	Met	His
				140					145					150
Glu	Arg	Arg	Ile	Leu	His	Arg	Asp	Leu	Lys	Ser	Lys	Asn	Val	Phe
				155					160					165
Leu	Lys	Asn	Asn	Leu	Leu	Lys	Ile	Gly	Asp	Phe	Gly	Val	Ser	Arg
				170					175					180
Leu	Leu	Met	Gly	Ser	Cys	Asp	Leu	Ala	Thr	Thr	Leu	Thr	Gly	Thr

Pro His Tyr Met	185	Pro Glu Ala Leu	190	Lys His Gln Gly Tyr Asp	195
	200		205		210
Thr Lys Ser Asp	215	Ile Trp Ser Leu Ala	220	Cys Ile Leu Tyr Glu Met	225
	230		235		240
Cys Cys Met Asn	245	His Ala Phe Ala Gly	250	Ser Asn Phe Leu Ser Ile	255
	260		265		270
Val Leu Lys Ile	275	Val Glu Gly Asp Thr	280	Pro Ser Leu Pro Glu Arg	285
	290		295		300
Tyr Pro Lys Glu	305	Leu Asn Ala Ile Met	310	Glu Ser Met Leu Asn Lys	315
	320		325		330
Asn Pro Ser Leu	335	Arg Pro Ser Ala Ile	340	Glu Ile Leu Lys Ile Pro	345
	350		355		360
Tyr Leu Asp Glu	365	Gln Leu Gln Asn Leu	370	Met Cys Arg Tyr Ser Glu	375
	380		385		390
Met Thr Leu Glu	395	Asp Lys Asn Leu Asp	400	Cys Gln Lys Glu Ala Ala	405
	410		415		420
His Ile Ile Asn	425	Ala Met Gln Lys Arg	430	Ile His Leu Gln Thr Leu	435
	440		445		450
Arg Ala Leu Ser	455	Glu Val Gln Lys Met	460	Thr Pro Arg Glu Arg Met	465
	470		475		480
Arg Leu Arg Lys	485	Leu Gln Ala Ala Asp	490	Glu Lys Ala Arg Lys Leu	495
	500		505		510
Lys Lys Ile Val	515	Glu Glu Lys Tyr Glu	520	Glu Asn Ser Lys Arg Met	525
	530		535		540
Gln Glu Leu Arg	545	Ser Arg Asn Phe Gln	550	Gln Leu Ser Val Asp Val	555
	560		565		570
Leu His Glu Lys	575	Thr His Leu Lys Gly	580	Met Glu Glu Lys Glu Glu	585
	590		595		600
Gln Pro Glu Gly	605	Arg Leu Ser Cys Ser	610	Pro Gln Asp Glu Asp Glu	615
	620		625		630
Glu Arg Trp Gln	635	Gly Arg Glu Glu Glu	640	Ser Asp Glu Pro Thr Leu	645
	650		655		660
Glu Asn Leu Pro	665	Glu Ser Gln Pro Ile	670	Pro Ser Met Asp Leu His	675
	680		685		690
Glu Leu Glu Ser	695	Ile Val Glu Asp Ala	700	Thr Ser Asp Leu Gly Tyr	705
	710		715		720
His Glu Ile Pro	725	Glu Asp Pro Leu Val	730	Ala Glu Glu Tyr Tyr Ala	735
	740		745		750
Asp Ala Phe Asp	755	Ser Tyr Cys Val Glu	760	Ser Asp Glu Glu Glu Glu	765
	770		775		780
Glu Ile Ala Leu	785	Glu Arg Pro Glu Lys	790	Glu Ile Arg Asn Glu Gly	795
	800		805		810
Ser Gln Pro Ala	815	Tyr Arg Thr Asn Gln	820	Gln Asp Ser Asp Ile Glu	825
	830		835		840
Ala Leu Ala Arg	845	Cys Leu Glu Asn Val	850	Leu Gly Cys Thr Ser Leu	855
	860		865		870
Asp Thr Lys Thr	875	Ile Thr Thr Met Ala	880	Glu Asp Met Ser Pro Gly	885
	890		895		900
Pro Pro Ile Phe	905	Asn Ser Val Met Ala	910	Arg Thr Lys Met Lys Arg	915
	920		925		930
Met Arg Glu Ser	935	Ala Met Gln Lys Leu	940	Gly Thr Glu Val Phe Glu	945
	950		955		960
Glu Val Tyr Asn	965	Tyr Leu Lys Arg Ala	970	Arg His Gln Asn Ala Ser	975
	980		985		990
Glu Ala Glu Ile	995	Arg Glu Cys Leu Glu	1000	Lys Val Val Pro Gln Ala	1005

	605		610		615
Ser Asp Cys Phe	Glu Val Asp Gln Leu	Leu Tyr Phe Glu Glu	Gln		
	620		625		630
Leu Leu Ile Thr Met Gly Lys Glu Pro	Thr Leu Gln Asn His	Leu			
	635		640		645

&lt;210&gt; 6

&lt;211&gt; 623

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477587CD1

&lt;400&gt; 6

Met Trp Ala Pro Gly Thr Arg Gln Gln Gly Gly Pro Glu Met Ala		
1	5	10 15
His Ile Gln Asn Val Glu Ala His Thr Ser Ser Ala Leu Trp Gly		
	20	25 30
Arg Ser Pro Arg Lys Pro Pro Thr Pro His Ala Arg Glu Ser Leu		
	35	40 45
Ser Phe Pro Leu Glu Arg Pro Arg Ser Gly Arg Ser Ala Val Val		
	50	55 60
Ser Ala Arg Leu Arg Gln Ser Pro Arg Met Glu Pro Arg Pro Arg		
	65	70 75
Arg Arg Arg Arg Ser Arg Pro Leu Val Ala Ala Phe Leu Arg Asp		
	80	85 90
Pro Gly Ser Gly Arg Val Tyr Arg Arg Gly Lys Leu Ile Gly Lys		
	95	100 105
Gly Ala Phe Ser Arg Cys Tyr Lys Leu Thr Asp Met Ser Thr Ser		
	110	115 120
Ala Val Phe Ala Leu Lys Val Val Pro Cys Gly Gly Ala Gly Ala		
	125	130 135
Gly Trp Leu Arg Pro Gln Gly Lys Val Glu Arg Glu Ile Ala Leu		
	140	145 150
His Ser Arg Leu Arg Pro Arg Asn Ile Val Ala Phe His Gly His		
	155	160 165
Phe Ala Asp Arg Asp His Val Tyr Met Val Leu Glu Tyr Cys Ser		
	170	175 180
Arg Gln Ser Leu Ala His Val Leu Arg Ala Arg Gln Ile Leu Thr		
	185	190 195
Glu Pro Glu Val Arg Asp Tyr Leu Arg Gly Leu Val Ser Gly Leu		
	200	205 210
Arg Tyr Leu His Gln Arg Cys Ile Leu His Arg Asp Leu Lys Leu		
	215	220 225
Ser Asn Phe Phe Leu Asn Lys Asn Met Glu Val Lys Ile Gly Asp		
	230	235 240
Leu Gly Leu Ala Ala Lys Val Gly Pro Gly Gly Arg Cys His Arg		
	245	250 255
Tyr Thr Val Leu Thr Gly Thr Pro Pro Phe Met Ala Ser Pro Leu		
	260	265 270
Ser Glu Met Tyr Gln Asn Ile Arg Glu Gly His Tyr Pro Glu Pro		
	275	280 285
Ala His Leu Ser Ala Asn Ala Arg Arg Leu Ile Val His Leu Leu		



290	295	300
Ala Pro Asn Pro	Ala Glu Arg Pro Ser	Leu Asp His Leu Leu Gln
305	310	315
Asp Asp Phe Phe	Thr Gln Gly Phe Thr	Pro Asp Arg Leu Pro Ala
320	325	330
His Ser Cys His	Ser Pro Pro Ile Phe	Ala Ile Pro Pro Pro Leu
335	340	345
Gly Arg Ile Phe	Arg Lys Val Gly Gln	Arg Leu Leu Thr Gln Cys
350	355	360
Arg Pro Pro Cys	Pro Phe Thr Pro Lys	Glu Ala Ser Gly Pro Gly
365	370	375
Glu Gly Gly Pro	Asp Pro Asp Ser Met	Glu Trp Asp Gly Glu Ser
380	385	390
Ser Leu Ser Ala	Lys Glu Val Pro Cys	Leu Glu Gly Pro Ile His
395	400	405
Leu Val Ala Gln	Gly Thr Leu Gln Ser	Asp Leu Ala Ala Thr Gln
410	415	420
Asp Pro Leu Gly	Glu Gln Gln Pro Ile	Leu Trp Ala Pro Lys Trp
425	430	435
Val Asp Tyr Ser	Ser Lys Tyr Gly Phe	Gly Tyr Gln Leu Leu Asp
440	445	450
Gly Gly Arg Thr	Gly Arg His Pro His	Gly Pro Ala Thr Pro Arg
455	460	465
Arg Tyr Leu Leu	Ser Thr Tyr Cys Ala	His Leu Gln Val Leu Pro
470	475	480
Ala Cys Gln Val	Cys Tyr Met Pro Asn	Cys Gly Arg Leu Glu Ala
485	490	495
Phe Ala Leu Arg	Asp Val Pro Gly Leu	Leu Gly Ala Lys Leu Ala
500	505	510
Val Leu Gln Leu	Phe Ala Gly Cys Leu	Arg Arg Arg Leu Arg Glu
515	520	525
Glu Gly Thr Leu	Pro Thr Pro Val Pro	Pro Ala Gly Pro Gly Leu
530	535	540
Cys Leu Leu Arg	Phe Leu Ala Ser Glu	His Ala Leu Leu Leu Leu
545	550	555
Phe Ser Asn Gly	Met Val Gln Val Ser	Phe Ser Gly Val Pro Ala
560	565	570
Gln Leu Val Leu	Ser Gly Glu Gly Glu	Gly Leu Gln Leu Thr Leu
575	580	585
Trp Glu Gln Gly	Ser Pro Gly Thr Ser	Tyr Ser Leu Asp Val Pro
590	595	600
Arg Ser His Gly	Cys Ala Pro Thr Thr	Gly Gln His Leu His His
605	610	615
Ala Leu Arg Met	Leu Gln Ser Ile	
620		

&lt;210&gt; 7

&lt;211&gt; 797

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7594537CD1

&lt;400&gt; 7

Met	Thr	Asn	Gln	Glu	Lys	Trp	Ala	His	Leu	Ser	Pro	Ser	Glu	Phe
1				5					10					15
Ser	Gln	Leu	Gln	Lys	Tyr	Ala	Glu	Tyr	Ser	Thr	Lys	Lys	Leu	Lys
				20					25					30
Asp	Val	Leu	Glu	Glu	Phe	His	Gly	Asn	Gly	Val	Leu	Ala	Lys	Tyr
				35					40					45
Asn	Pro	Glu	Gly	Thr	Ile	Asp	Phe	Glu	Gly	Phe	Lys	Leu	Phe	Met
				50					55					60
Lys	Thr	Phe	Leu	Glu	Ala	Glu	Leu	Pro	Asp	Asp	Phe	Thr	Ala	His
				65					70					75
Leu	Phe	Met	Ser	Phe	Ser	Asn	Lys	Phe	Pro	His	Ser	Ser	Pro	Met
				80					85					90
Val	Lys	Ser	Lys	Pro	Ala	Leu	Leu	Ser	Gly	Gly	Leu	Arg	Met	Asn
				95					100					105
Lys	Gly	Ala	Ile	Thr	Pro	Pro	Arg	Thr	Thr	Ser	Pro	Ala	Asn	Thr
				110					115					120
Cys	Ser	Pro	Glu	Val	Ile	His	Leu	Lys	Asp	Ile	Val	Cys	Tyr	Leu
				125					130					135
Ser	Leu	Leu	Glu	Arg	Gly	Arg	Pro	Glu	Asp	Lys	Leu	Glu	Phe	Met
				140					145					150
Phe	Arg	Leu	Tyr	Asp	Thr	Asp	Gly	Asn	Gly	Phe	Leu	Asp	Ser	Ser
				155					160					165
Glu	Leu	Glu	Asn	Ile	Ile	Ser	Gln	Met	Met	His	Val	Ala	Glu	Tyr
				170					175					180
Leu	Glu	Trp	Asp	Val	Thr	Glu	Leu	Asn	Pro	Ile	Leu	His	Glu	Met
				185					190					195
Met	Glu	Glu	Ile	Asp	Tyr	Asp	His	Asp	Gly	Thr	Val	Ser	Leu	Glu
				200					205					210
Glu	Trp	Ile	Gln	Gly	Gly	Met	Thr	Thr	Ile	Pro	Leu	Leu	Val	Leu
				215					220					225
Leu	Gly	Leu	Glu	Asn	Asn	Val	Lys	Asp	Asp	Gly	Gln	His	Val	Trp
				230					235					240
Arg	Leu	Lys	His	Phe	Asn	Lys	Pro	Ala	Tyr	Cys	Asn	Leu	Cys	Leu
				245					250					255
Asn	Met	Leu	Ile	Gly	Val	Gly	Lys	Gln	Gly	Leu	Cys	Cys	Ser	Phe
				260					265					270
Cys	Lys	Tyr	Thr	Val	His	Glu	Arg	Cys	Val	Ala	Arg	Ala	Pro	Pro
				275					280					285
Ser	Cys	Ile	Lys	Thr	Tyr	Val	Lys	Ser	Lys	Arg	Asn	Thr	Asp	Val
				290					295					300
Met	His	His	Tyr	Trp	Val	Glu	Gly	Asn	Cys	Pro	Thr	Lys	Cys	Asp
				305					310					315
Lys	Cys	His	Lys	Thr	Val	Lys	Cys	Tyr	Gln	Gly	Leu	Thr	Gly	Leu
				320					325					330
His	Cys	Val	Trp	Cys	Gln	Ile	Thr	Leu	His	Asn	Lys	Cys	Ala	Ser
				335					340					345
His	Leu	Lys	Pro	Glu	Cys	Asp	Cys	Gly	Pro	Leu	Lys	Asp	His	Ile
				350					355					360
Leu	Pro	Pro	Thr	Thr	Ile	Cys	Pro	Val	Val	Leu	Gln	Thr	Leu	Pro
				365					370					375
Thr	Ser	Gly	Val	Ser	Val	Pro	Glu	Glu	Arg	Gln	Ser	Thr	Val	Lys
				380					385					390
Lys	Glu	Lys	Ser	Gly	Ser	Gln	Gln	Pro	Asn	Lys	Val	Ile	Asp	Lys
				395					400					405
Asn	Lys	Met	Gln	Arg	Ala	Asn	Ser	Val	Thr	Val	Asp	Gly	Gln	Gly
				410					415					420

Leu	Gln	Val	Thr	Pro	Val	Pro	Gly	Thr	His	Pro	Leu	Leu	Val	Phe
				425					430					435
Val	Asn	Pro	Lys	Ser	Gly	Gly	Lys	Gln	Gly	Glu	Arg	Ile	Tyr	Arg
				440					445					450
Lys	Phe	Gln	Tyr	Leu	Leu	Asn	Pro	Arg	Gln	Val	Tyr	Ser	Leu	Ser
				455					460					465
Gly	Asn	Gly	Pro	Met	Pro	Gly	Leu	Asn	Phe	Phe	Arg	Asp	Val	Pro
				470					475					480
Asp	Phe	Arg	Val	Leu	Ala	Cys	Gly	Gly	Asp	Gly	Thr	Val	Gly	Trp
				485					490					495
Val	Leu	Asp	Cys	Ile	Glu	Lys	Ala	Asn	Val	Gly	Lys	His	Pro	Pro
				500					505					510
Val	Ala	Ile	Leu	Pro	Leu	Gly	Thr	Gly	Asn	Asp	Leu	Ala	Arg	Cys
				515					520					525
Leu	Arg	Trp	Gly	Gly	Gly	Tyr	Glu	Gly	Glu	Asn	Leu	Met	Lys	Ile
				530					535					540
Leu	Lys	Asp	Ile	Glu	Asn	Ser	Thr	Glu	Ile	Met	Leu	Asp	Arg	Trp
				545					550					555
Lys	Phe	Glu	Val	Ile	Pro	Asn	Asp	Lys	Asp	Glu	Lys	Gly	Asp	Pro
				560					565					570
Val	Pro	Tyr	Ser	Ile	Ile	Asn	Asn	Tyr	Phe	Ser	Ile	Gly	Val	Asp
				575					580					585
Ala	Ser	Ile	Ala	His	Arg	Phe	His	Ile	Met	Arg	Glu	Lys	His	Pro
				590					595					600
Glu	Lys	Phe	Asn	Ser	Arg	Met	Lys	Asn	Lys	Phe	Trp	Tyr	Phe	Glu
				605					610					615
Phe	Gly	Thr	Ser	Glu	Thr	Phe	Ser	Ala	Thr	Cys	Lys	Lys	Leu	His
				620					625					630
Glu	Ser	Val	Glu	Ile	Glu	Cys	Asp	Gly	Val	Gln	Ile	Asp	Leu	Ile
				635					640					645
Asn	Ile	Ser	Leu	Glu	Gly	Ile	Ala	Ile	Leu	Asn	Ile	Pro	Ser	Met
				650					655					660
His	Gly	Gly	Ser	Asn	Leu	Trp	Gly	Glu	Ser	Lys	Lys	Arg	Arg	Ser
				665					670					675
His	Arg	Arg	Ile	Glu	Lys	Lys	Gly	Ser	Asp	Lys	Arg	Thr	Thr	Val
				680					685					690
Thr	Asp	Ala	Lys	Glu	Leu	Lys	Phe	Ala	Ser	Gln	Asp	Leu	Ser	Asp
				695					700					705
Gln	Leu	Leu	Glu	Val	Val	Gly	Leu	Glu	Gly	Ala	Met	Glu	Met	Gly
				710					715					720
Gln	Ile	Tyr	Thr	Gly	Leu	Lys	Ser	Ala	Gly	Arg	Arg	Leu	Ala	Gln
				725					730					735
Cys	Ser	Cys	Val	Val	Ile	Arg	Thr	Ser	Lys	Ser	Leu	Pro	Met	Gln
				740					745					750
Ile	Asp	Gly	Glu	Pro	Trp	Met	Gln	Thr	Pro	Cys	Thr	Ile	Lys	Ile
				755					760					765
Thr	His	Lys	Asn	Gln	Ala	Pro	Met	Leu	Met	Gly	Pro	Pro	Pro	Lys
				770					775					780
Thr	Gly	Leu	Phe	Cys	Ser	Leu	Val	Lys	Arg	Thr	Arg	Asn	Arg	Ser
				785					790					795
Lys	Glu													

&lt;210&gt; 8

&lt;211&gt; 749

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70467491CD1

&lt;400&gt; 8

Met	Ser	Thr	Arg	Thr	Pro	Leu	Pro	Thr	Val	Asn	Glu	Arg	Asp	Thr	1	5	10	15
Glu	Asn	Ala	Val	Leu	Pro	His	Thr	Ser	His	Gly	Asp	Gly	Arg	Gln	20	25	30	35
Glu	Val	Thr	Ser	Arg	Thr	Ser	Arg	Ser	Gly	Ala	Arg	Cys	Arg	Asn	35	40	45	50
Ser	Ile	Ala	Ser	Cys	Ala	Asp	Glu	Gln	Pro	His	Ile	Gly	Asn	Tyr	50	55	60	65
Arg	Leu	Leu	Lys	Thr	Ile	Gly	Lys	Gly	Asn	Phe	Ala	Lys	Val	Lys	65	70	75	80
Leu	Ala	Arg	His	Ile	Leu	Thr	Gly	Arg	Glu	Lys	Asn	Val	Arg	Ile	80	85	90	95
Ser	Lys	Glu	Ile	Asp	Asn	Phe	Leu	Gly	Lys	His	Asp	Leu	Pro	Lys	95	100	105	110
Leu	Thr	Leu	Glu	Lys	Asn	Arg	Tyr	Thr	Ser	Val	Thr	Thr	Glu	Val	110	115	120	125
Glu	Lys	Val	Val	Asn	Ile	Leu	Pro	Asn	Leu	Glu	Phe	Met	Ile	Glu	125	130	135	140
Phe	Phe	Glu	Ile	Tyr	Ser	Ile	Gly	Glu	Val	Phe	Asp	Tyr	Leu	Val	140	145	150	155
Ala	His	Gly	Arg	Met	Lys	Glu	Lys	Glu	Ala	Arg	Ser	Lys	Phe	Arg	155	160	165	170
Gln	Ile	Val	Ser	Ala	Val	Gln	Tyr	Cys	His	Gln	Lys	Arg	Ile	Val	170	175	180	185
His	Arg	Asp	Leu	Lys	Ala	Glu	Asn	Leu	Leu	Leu	Asp	Ala	Asp	Met	185	190	195	200
Asn	Ile	Lys	Ile	Ala	Asp	Phe	Gly	Phe	Ser	Asn	Glu	Phe	Thr	Val	200	205	210	215
Gly	Gly	Lys	Leu	Asp	Thr	Phe	Cys	Gly	Ser	Pro	Pro	Tyr	Ala	Ala	215	220	225	230
Pro	Glu	Leu	Phe	Gln	Gly	Lys	Lys	Tyr	Asp	Gly	Pro	Glu	Val	Asp	230	235	240	245
Val	Trp	Ser	Leu	Gly	Val	Ile	Leu	Tyr	Thr	Leu	Val	Ser	Gly	Ser	245	250	255	260
Leu	Pro	Phe	Asp	Gly	Gln	Asn	Leu	Lys	Glu	Leu	Arg	Glu	Arg	Val	260	265	270	275
Leu	Arg	Gly	Lys	Tyr	Arg	Ile	Pro	Phe	Tyr	Met	Ser	Thr	Asp	Cys	275	280	285	290
Glu	Asn	Leu	Leu	Lys	Arg	Phe	Leu	Val	Leu	Asn	Pro	Ile	Lys	Arg	290	295	300	305
Gly	Thr	Leu	Glu	Gln	Ile	Met	Lys	Asp	Arg	Trp	Ile	Asn	Ala	Gly	305	310	315	320
His	Glu	Glu	Asp	Glu	Leu	Lys	Pro	Phe	Val	Glu	Pro	Glu	Leu	Asp	320	325	330	335
Ile	Ser	Asp	Gln	Lys	Arg	Ile	Asp	Ile	Met	Val	Gly	Met	Gly	Tyr	335	340	345	350
Ser	Gln	Glu	Glu	Ile	Gln	Glu	Ser	Leu	Ser	Lys	Met	Lys	Tyr	Asp	350	355	360	
Glu	Ile	Thr	Ala	Thr	Tyr	Leu	Leu	Leu	Gly	Arg	Lys	Ser	Ser	Glu				

	365	370	375
Leu Asp Ala Ser	Asp Ser Ser Ser Ser	Ser Asn Leu Ser	Leu Ala
	380	385	390
Lys Val Arg Pro	Ser Ser Asp Leu Asn	Asn Ser Thr Gly	Gln Ser
	395	400	405
Pro His His Lys	Val Gln Arg Ser Val	Ser Ser Ser Gln	Lys Gln
	410	415	420
Arg Arg Tyr Ser	Asp His Ala Gly Pro	Ala Ile Pro Ser	Val Val
	425	430	435
Ala Tyr Pro Lys	Arg Ser Gln Thr Ser	Thr Ala Asp Ser	Asp Leu
	440	445	450
Lys Glu Asp Gly	Ile Ser Ser Arg Lys	Ser Ser Gly Ser	Ala Val
	455	460	465
Gly Gly Lys Gly	Ile Ala Pro Ala Ser	Pro Met Leu Gly	Asn Ala
	470	475	480
Ser Asn Pro Asn	Lys Ala Asp Ile Pro	Glu Arg Lys Lys	Ser Ser
	485	490	495
Thr Val Pro Ser	Ser Asn Thr Ala Ser	Gly Gly Met Thr	Arg Arg
	500	505	510
Asn Thr Tyr Val	Cys Ser Glu Arg Thr	Thr Ala Asp Arg	His Ser
	515	520	525
Val Ile Gln Asn	Gly Lys Glu Asn Ser	Thr Ile Pro Asp	Gln Arg
	530	535	540
Thr Pro Val Ala	Ser Thr His Ser Ile	Ser Ser Ala Ala	Thr Pro
	545	550	555
Asp Arg Ile Arg	Phe Pro Arg Gly Thr	Ala Ser Arg Ser	Thr Phe
	560	565	570
His Gly Gln Pro	Arg Glu Arg Arg Thr	Ala Thr Tyr Asn	Gly Pro
	575	580	585
Pro Ala Ser Pro	Ser Leu Ser His Glu	Ala Thr Pro Leu	Ser Gln
	590	595	600
Thr Arg Ser Arg	Gly Ser Thr Asn Leu	Phe Ser Lys Leu	Thr Ser
	605	610	615
Lys Leu Thr Arg	Arg Leu Pro Thr Glu	Tyr Glu Arg Asn	Gly Arg
	620	625	630
Tyr Glu Gly Ser	Ser Arg Asn Val Ser	Ala Glu Gln Lys	Asp Glu
	635	640	645
Asn Lys Glu Ala	Lys Pro Arg Ser Leu	Arg Phe Thr Trp	Ser Met
	650	655	660
Lys Thr Thr Ser	Ser Met Asp Pro Gly	Asp Met Met Arg	Glu Ile
	665	670	675
Arg Lys Val Leu	Asp Ala Asn Asn Cys	Asp Tyr Glu Gln	Arg Glu
	680	685	690
Arg Phe Leu Leu	Phe Cys Val His Gly	Asp Gly His Ala	Glu Asn
	695	700	705
Leu Val Gln Trp	Glu Met Glu Val Cys	Lys Leu Pro Arg	Leu Ser
	710	715	720
Leu Asn Gly Val	Arg Phe Lys Arg Ile	Ser Gly Thr Ser	Ile Ala
	725	730	735
Phe Lys Asn Ile	Ala Ser Lys Ile Ala	Asn Glu Leu Lys	Leu
	740	745	

&lt;210&gt; 9

&lt;211&gt; 386

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478559CD1

&lt;400&gt; 9

```

Met Ala Val Pro Pro Ser Ala Pro Gln Pro Arg Ala Ser Phe His
 1          5          10          15
Leu Arg Arg His Thr Pro Cys Pro Gln Cys Ser Trp Gly Met Glu
          20          25          30
Glu Lys Ala Ala Ala Ser Ala Ser Cys Arg Glu Pro Pro Gly Pro
          35          40          45
Pro Arg Ala Ala Ala Val Ala Tyr Phe Gly Ile Ser Val Asp Pro
          50          55          60
Asp Asp Ile Leu Pro Gly Ala Leu Arg Leu Ile Gln Glu Leu Arg
          65          70          75
Pro His Trp Lys Pro Glu Gln Val Arg Thr Lys Arg Phe Met Asp
          80          85          90
Gly Ile Thr Asn Lys Leu Val Ala Cys Tyr Val Glu Glu Asp Met
          95          100          105
Gln Asp Cys Val Leu Val Arg Val Tyr Gly Glu Arg Thr Glu Leu
          110          115          120
Leu Val Asp Arg Glu Asn Glu Val Arg Asn Phe Gln Leu Leu Arg
          125          130          135
Ala His Ser Cys Ala Pro Lys Leu Tyr Cys Thr Phe Gln Asn Gly
          140          145          150
Leu Cys Tyr Glu Tyr Met Gln Gly Val Ala Leu Glu Pro Glu His
          155          160          165
Ile Arg Glu Pro Arg Leu Phe Arg Leu Ile Ala Leu Glu Met Ala
          170          175          180
Lys Ile His Thr Ile His Ala Asn Gly Ser Leu Pro Lys Pro Ile
          185          190          195
Leu Trp His Lys Met His Asn Tyr Phe Thr Leu Val Lys Asn Glu
          200          205          210
Ile Asn Pro Ser Leu Ser Ala Asp Val Pro Lys Val Glu Val Leu
          215          220          225
Glu Arg Glu Leu Ala Trp Leu Lys Glu His Leu Ser Gln Leu Glu
          230          235          240
Ser Pro Val Val Phe Cys His Asn Asp Leu Leu Cys Lys Asn Ile
          245          250          255
Ile Tyr Asp Ser Ile Lys Gly His Val Arg Phe Ile Asp Tyr Glu
          260          265          270
Tyr Ala Gly Tyr Asn Tyr Gln Ala Phe Asp Ile Gly Asn His Phe
          275          280          285
Asn Glu Phe Ala Gly Val Asn Glu Val Asp Tyr Cys Leu Tyr Pro
          290          295          300
Ala Arg Glu Thr Gln Leu Gln Trp Leu His Tyr Tyr Leu Gln Ala
          305          310          315
Gln Lys Gly Met Ala Val Thr Pro Arg Glu Val Gln Arg Leu Tyr
          320          325          330
Val Gln Val Asn Lys Phe Ala Leu Ala Ser His Phe Phe Trp Ala
          335          340          345
Leu Trp Ala Leu Ile Gln Asn Gln Tyr Ser Thr Ile Asp Phe Asp
          350          355          360
Phe Leu Arg Tyr Ala Val Ile Arg Phe Asn Gln Tyr Phe Lys Val
          365          370          375
Lys Pro Gln Ala Ser Ala Leu Glu Met Pro Lys

```

380

385

<210> 10  
 <211> 342  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1698381CD1

&lt;400&gt; 10

Met	Glu	Lys	Tyr	Glu	Lys	Leu	Ala	Lys	Thr	Gly	Glu	Gly	Ser	Tyr	1	5	10	15
Gly	Val	Val	Phe	Lys	Cys	Arg	Asn	Lys	Thr	Ser	Gly	Gln	Val	Val	20	25	30	
Ala	Val	Lys	Lys	Phe	Val	Glu	Ser	Glu	Asp	Asp	Pro	Val	Val	Lys	35	40	45	
Lys	Ile	Ala	Leu	Arg	Glu	Ile	Arg	Met	Leu	Lys	Gln	Leu	Lys	His	50	55	60	
Pro	Asn	Leu	Val	Asn	Leu	Ile	Glu	Val	Phe	Arg	Arg	Lys	Arg	Lys	65	70	75	
Met	His	Leu	Val	Phe	Glu	Tyr	Cys	Asp	His	Thr	Leu	Leu	Asn	Glu	80	85	90	
Leu	Glu	Arg	Asn	Pro	Asn	Gly	Val	Ala	Asp	Gly	Val	Ile	Lys	Ser	95	100	105	
Val	Leu	Trp	Gln	Thr	Leu	Gln	Ala	Leu	Asn	Phe	Cys	His	Ile	His	110	115	120	
Asn	Cys	Ile	His	Arg	Asp	Ile	Lys	Pro	Glu	Asn	Ile	Leu	Ile	Thr	125	130	135	
Lys	Gln	Gly	Ile	Ile	Lys	Ile	Cys	Asp	Phe	Gly	Phe	Ala	Gln	Ile	140	145	150	
Leu	Ile	Pro	Gly	Asp	Ala	Tyr	Thr	Asp	Tyr	Val	Ala	Thr	Arg	Trp	155	160	165	
Tyr	Arg	Ala	Pro	Glu	Leu	Leu	Val	Gly	Asp	Thr	Gln	Tyr	Gly	Ser	170	175	180	
Ser	Val	Asp	Ile	Trp	Ala	Ile	Gly	Cys	Val	Phe	Ala	Glu	Leu	Leu	185	190	195	
Thr	Gly	Gln	Pro	Leu	Trp	Pro	Gly	Lys	Ser	Asp	Val	Asp	Gln	Leu	200	205	210	
Tyr	Leu	Ile	Ile	Arg	Thr	Leu	Gly	Lys	Leu	Ile	Pro	Arg	His	Gln	215	220	225	
Ser	Ile	Phe	Lys	Ser	Asn	Gly	Phe	Phe	His	Gly	Ile	Ser	Ile	Pro	230	235	240	
Glu	Pro	Glu	Asp	Met	Glu	Thr	Leu	Glu	Glu	Lys	Phe	Ser	Asp	Val	245	250	255	
His	Pro	Val	Ala	Leu	Asn	Phe	Met	Lys	Gly	Cys	Leu	Lys	Met	Asn	260	265	270	
Pro	Asp	Asp	Arg	Leu	Thr	Cys	Ser	Gln	Leu	Leu	Glu	Ser	Ser	Tyr	275	280	285	
Phe	Asp	Ser	Phe	Gln	Glu	Ala	Gln	Ile	Lys	Arg	Lys	Ala	Arg	Asn	290	295	300	
Glu	Gly	Arg	Asn	Arg	Arg	Gln	Gln	Gln	Asn	Gln	Leu	Leu	Pro	Leu	305	310	315	
Ile	Pro	Gly	Ser	His	Ile	Ser	Pro	Thr	Pro	Asp	Gly	Arg	Lys	Gln	320	325	330	

Val Leu Gln Leu Lys Phe Asp His Leu Pro Asn Ile  
335 340

<210> 11

<211> 1164

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474637CD1

<400> 11

Met	Ala	Gly	Ala	Gly	Gly	Gln	His	His	Pro	Pro	Gly	Ala	Ala	Gly	1	5	10	15
Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Ala	Val	Thr	Ser	Ala	Ala	Ala	20	25	30	
Ser	Ala	Gly	Pro	Gly	Glu	Asp	Ser	Ser	Asp	Ser	Glu	Ala	Glu	Gln	35	40	45	
Glu	Gly	Pro	Gln	Lys	Leu	Ile	Arg	Lys	Val	Ser	Thr	Ser	Gly	Gln	50	55	60	
Ile	Arg	Thr	Lys	Thr	Ser	Ile	Lys	Glu	Gly	Gln	Leu	Leu	Lys	Gln	65	70	75	
Thr	Ser	Ser	Phe	Gln	Arg	Trp	Lys	Lys	Arg	Tyr	Phe	Lys	Leu	Arg	80	85	90	
Gly	Arg	Thr	Leu	Tyr	Tyr	Ala	Lys	Asp	Ser	Lys	Ser	Leu	Ile	Phe	95	100	105	
Asp	Glu	Val	Asp	Leu	Ser	Asp	Ala	Ser	Val	Ala	Glu	Ala	Ser	Thr	110	115	120	
Lys	Asn	Ala	Asn	Asn	Ser	Phe	Thr	Ile	Ile	Thr	Pro	Phe	Arg	Arg	125	130	135	
Leu	Met	Leu	Cys	Ala	Glu	Asn	Arg	Lys	Glu	Met	Glu	Asp	Trp	Ile	140	145	150	
Ser	Ser	Leu	Lys	Ser	Val	Gln	Thr	Arg	Glu	Pro	Tyr	Glu	Val	Ala	155	160	165	
Gln	Phe	Asn	Val	Glu	His	Phe	Ser	Gly	Met	His	Asn	Trp	Tyr	Ala	170	175	180	
Cys	Ser	His	Ala	Arg	Pro	Thr	Phe	Cys	Asn	Val	Cys	Arg	Glu	Ser	185	190	195	
Leu	Ser	Gly	Val	Thr	Ser	His	Gly	Leu	Ser	Cys	Glu	Val	Cys	Lys	200	205	210	
Phe	Lys	Ala	His	Lys	Arg	Cys	Ala	Val	Arg	Ala	Thr	Asn	Asn	Cys	215	220	225	
Lys	Trp	Thr	Thr	Leu	Ala	Ser	Ile	Gly	Lys	Asp	Ile	Ile	Glu	Asp	230	235	240	
Glu	Asp	Gly	Val	Ala	Met	Pro	His	Gln	Trp	Leu	Glu	Gly	Asn	Leu	245	250	255	
Pro	Val	Ser	Ala	Lys	Cys	Ala	Val	Cys	Asp	Lys	Thr	Cys	Gly	Ser	260	265	270	
Val	Leu	Arg	Leu	Gln	Asp	Trp	Lys	Cys	Leu	Trp	Cys	Lys	Thr	Met	275	280	285	
Val	His	Thr	Ala	Cys	Lys	Asp	Leu	Tyr	His	Pro	Ile	Cys	Pro	Leu	290	295	300	
Gly	Gln	Cys	Lys	Val	Ser	Ile	Ile	Pro	Pro	Ile	Ala	Leu	Asn	Ser	305	310	315	
Thr	Asp	Ser	Asp	Gly	Phe	Cys	Arg	Ala	Thr	Phe	Ser	Phe	Cys	Val				



320	325	330
Ser Pro Leu Leu Val Phe Val Asn Ser	Lys Ser Gly Asp Asn Gln	
335	340	345
Gly Val Lys Phe Leu Arg Arg Phe Lys	Gln Leu Leu Asn Pro Ala	
350	355	360
Gln Val Phe Asp Leu Met Asn Gly Gly	Pro His Leu Gly Leu Arg	
365	370	375
Leu Phe Gln Lys Phe Asp Asn Phe Arg	Ile Leu Val Cys Gly Gly	
380	385	390
Asp Gly Ser Val Gly Trp Val Leu Ser	Glu Ile Asp Lys Leu Asn	
395	400	405
Leu Asn Lys Gln Cys Gln Leu Gly Val	Leu Pro Leu Gly Thr Gly	
410	415	420
Asn Asp Leu Ala Arg Val Leu Gly Trp	Gly Gly Ser Tyr Asp Asp	
425	430	435
Asp Thr Gln Leu Pro Gln Ile Leu Glu	Lys Leu Glu Arg Ala Ser	
440	445	450
Thr Lys Met Leu Asp Arg Trp Ser Ile	Met Thr Tyr Glu Leu Lys	
455	460	465
Leu Pro Pro Lys Ala Ser Leu Leu Pro	Gly Pro Pro Glu Ala Ser	
470	475	480
Glu Glu Phe Tyr Met Thr Ile Tyr Glu	Asp Ser Val Ala Thr His	
485	490	495
Leu Thr Lys Ile Leu Asn Ser Asp Glu	His Ala Val Val Ile Ser	
500	505	510
Ser Ala Lys Thr Leu Cys Glu Thr Val	Lys Asp Phe Val Ala Lys	
515	520	525
Val Glu Lys Thr Tyr Asp Lys Thr Leu	Glu Asn Ala Val Val Ala	
530	535	540
Asp Ala Val Ala Ser Lys Cys Ser Val	Leu Asn Glu Lys Leu Glu	
545	550	555
Gln Leu Leu Gln Ala Leu His Thr Asp	Ser Gln Ala Ala Pro Val	
560	565	570
Leu Pro Gly Leu Ser Pro Leu Ile Val	Glu Glu Asp Ala Val Glu	
575	580	585
Ser Ser Ser Glu Glu Ser Leu Gly Glu	Ser Lys Glu Gln Leu Gly	
590	595	600
Asp Asp Val Thr Lys Pro Ser Ser Gln	Lys Ala Val Lys Pro Arg	
605	610	615
Glu Ile Met Leu Arg Ala Asn Ser Leu	Lys Lys Ala Val Arg Gln	
620	625	630
Val Ile Glu Glu Ala Gly Lys Val Met	Asp Asp Pro Thr Val His	
635	640	645
Pro Cys Glu Pro Ala Asn Gln Ser Ser	Asp Tyr Asp Ser Thr Glu	
650	655	660
Thr Asp Glu Ser Lys Glu Glu Ala Lys	Asp Asp Gly Ala Lys Glu	
665	670	675
Ser Ile Thr Val Lys Thr Ala Pro Arg	Ser Pro Asp Ala Arg Ala	
680	685	690
Ser Tyr Gly His Ser Gln Thr Asp Ser	Val Pro Gly Pro Ala Val	
695	700	705
Ala Ala Ser Lys Glu Asn Leu Pro Val	Leu Asn Thr Arg Ile Ile	
710	715	720
Cys Pro Gly Leu Arg Ala Gly Leu Ala	Ala Ser Ile Ala Gly Ser	
725	730	735
Ser Ile Ile Asn Lys Met Leu Leu Ala	Asn Ile Asp Pro Phe Gly	

740	745	750
Ala Thr Pro Phe Ile Asp Pro Asp Leu Asp Ser Val Asp Gly Tyr		
755	760	765
Ser Glu Lys Cys Val Met Asn Asn Tyr Phe Gly Ile Gly Leu Asp		
770	775	780
Ala Lys Ile Ser Leu Glu Phe Asn Asn Lys Arg Glu Glu His Pro		
785	790	795
Glu Lys Cys Arg Ser Arg Thr Lys Asn Leu Met Trp Tyr Gly Val		
800	805	810
Leu Gly Thr Arg Glu Leu Leu Gln Arg Ser Tyr Lys Asn Leu Glu		
815	820	825
Gln Arg Val Gln Leu Glu Cys Asp Gly Gln Tyr Ile Pro Leu Pro		
830	835	840
Ser Leu Gln Gly Ile Ala Val Leu Asn Ile Pro Ser Tyr Ala Gly		
845	850	855
Gly Thr Asn Phe Trp Gly Gly Thr Lys Glu Asp Asp Ile Phe Ala		
860	865	870
Ala Pro Ser Phe Asp Asp Lys Ile Leu Glu Val Val Ala Ile Phe		
875	880	885
Asp Ser Met Gln Met Ala Val Ser Arg Val Ile Lys Leu Gln His		
890	895	900
His Arg Ile Ala Gln Cys Arg Thr Val Lys Ile Thr Ile Phe Gly		
905	910	915
Asp Glu Gly Val Pro Val Gln Val Asp Gly Glu Ala Trp Val Gln		
920	925	930
Pro Pro Gly Ile Ile Lys Ile Val His Lys Asn Arg Ala Gln Met		
935	940	945
Leu Thr Arg Asp Arg Ala Phe Glu Ser Thr Leu Lys Ser Trp Glu		
950	955	960
Asp Lys Gln Lys Cys Asp Ser Gly Lys Pro Val Leu Arg Thr His		
965	970	975
Leu Tyr Ile His His Ala Ile Asp Leu Ala Thr Glu Glu Val Ser		
980	985	990
Gln Met Gln Leu Cys Ser Gln Ala Ala Glu Glu Leu Ile Thr Arg		
995	1000	1005
Ile Cys Asp Ala Ala Thr Ile His Cys Leu Leu Glu Gln Glu Leu		
1010	1015	1020
Ala His Ala Val Asn Ala Cys Ser His Ala Leu Asn Lys Ala Asn		
1025	1030	1035
Pro Arg Cys Pro Glu Ser Leu Thr Arg Asp Thr Ala Thr Glu Ile		
1040	1045	1050
Ala Ile Asn Val Lys Ala Leu Tyr Asn Glu Thr Glu Ser Leu Leu		
1055	1060	1065
Val Gly Arg Val Pro Leu Gln Leu Glu Ser Pro His Glu Glu Arg		
1070	1075	1080
Val Ser Asn Ala Leu His Ser Val Glu Val Glu Leu Gln Lys Leu		
1085	1090	1095
Thr Glu Ile Pro Trp Leu Tyr Tyr Ile Leu His Pro Asn Glu Asp		
1100	1105	1110
Glu Glu Pro Pro Met Asp Cys Thr Lys Arg Asn Asn Arg Ser Thr		
1115	1120	1125
Val Phe Arg Ile Val Pro Lys Phe Lys Lys Glu Lys Val Gln Lys		
1130	1135	1140
Gln Lys Thr Ser Ser Gln Pro Gly Ser Gly Asp Thr Glu Ser Gly		
1145	1150	1155
Ser Cys Glu Ala Asn Ser Pro Gly Asn		

1160

<210> 12  
 <211> 268  
 <212> PRT  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7170260CD1

<400> 12  
 Met Glu Asp Phe Leu Leu Ser Asn Gly Tyr Gln Leu Gly Lys Thr  
 1 5 10 15  
 Ile Gly Glu Gly Thr Tyr Ser Lys Val Lys Glu Ala Phe Ser Lys  
 20 25 30  
 Lys His Gln Arg Lys Val Ala Ile Lys Val Ile Asp Lys Met Gly  
 35 40 45  
 Gly Pro Glu Glu Phe Ile Gln Arg Phe Leu Pro Arg Glu Leu Gln  
 50 55 60  
 Ile Val Arg Thr Leu Asp His Lys Asn Ile Ile Gln Val Tyr Glu  
 65 70 75  
 Met Leu Glu Ser Ala Asp Gly Lys Ile Cys Leu Val Met Glu Leu  
 80 85 90  
 Ala Glu Gly Gly Asp Val Phe Asp Cys Val Leu Asn Gly Gly Pro  
 95 100 105  
 Leu Pro Glu Ser Arg Ala Lys Ala Leu Phe Arg Gln Met Val Glu  
 110 115 120  
 Ala Ile Arg Tyr Cys His Gly Cys Gly Val Ala His Arg Asp Leu  
 125 130 135  
 Lys Cys Glu Asn Ala Leu Leu Gln Gly Phe Asn Leu Lys Leu Thr  
 140 145 150  
 Asp Phe Gly Phe Ala Lys Val Leu Pro Lys Ser His Arg Glu Leu  
 155 160 165  
 Ser Gln Thr Phe Cys Gly Ser Thr Ala Tyr Ala Ala Pro Glu Val  
 170 175 180  
 Leu Gln Gly Ile Pro His Asp Ser Lys Lys Gly Asp Val Trp Ser  
 185 190 195  
 Met Gly Val Val Leu Tyr Val Met Leu Cys Ala Ser Leu Pro Phe  
 200 205 210  
 Asp Asp Thr Asp Ile Pro Lys Met Leu Trp Gln Gln Gln Lys Gly  
 215 220 225  
 Val Ser Phe Pro Thr His Leu Ser Ile Ser Ala Asp Cys Gln Asp  
 230 235 240  
 Leu Leu Lys Arg Leu Leu Glu Pro Asp Met Ile Leu Arg Pro Ser  
 245 250 255  
 Ile Glu Glu Val Ser Trp His Pro Trp Leu Ala Ser Thr  
 260 265

<210> 13  
 <211> 965  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 1797506CD1

&lt;400&gt; 13

Met	Arg	Arg	Ala	Gly	Ile	Gly	Glu	Asp	Ser	Arg	Leu	Gly	Leu	Gln
1				5					10					15
Ala	Gln	Pro	Gly	Ala	Glu	Pro	Ser	Pro	Gly	Arg	Ala	Gly	Thr	Glu
			20						25					30
Arg	Ser	Leu	Gly	Gly	Thr	Gln	Gly	Pro	Gly	Gln	Pro	Cys	Ser	Cys
			35						40					45
Pro	Gly	Ala	Met	Ala	Ser	Ala	Val	Arg	Gly	Ser	Arg	Pro	Trp	Pro
			50						55					60
Arg	Leu	Gly	Leu	Gln	Leu	Gln	Phe	Ala	Ala	Leu	Leu	Leu	Gly	Thr
			65						70					75
Leu	Ser	Pro	Gln	Val	His	Thr	Leu	Arg	Pro	Glu	Asn	Leu	Leu	Leu
			80						85					90
Val	Ser	Thr	Leu	Asp	Gly	Ser	Leu	His	Ala	Leu	Ser	Lys	Gln	Thr
			95						100					105
Gly	Asp	Leu	Lys	Trp	Thr	Leu	Arg	Asp	Asp	Pro	Val	Ile	Glu	Gly
			110						115					120
Pro	Met	Tyr	Val	Thr	Glu	Met	Ala	Phe	Leu	Ser	Asp	Pro	Ala	Asp
			125						130					135
Gly	Ser	Leu	Tyr	Ile	Leu	Gly	Thr	Gln	Lys	Gln	Gln	Gly	Leu	Met
			140						145					150
Lys	Leu	Pro	Phe	Thr	Ile	Pro	Glu	Leu	Val	His	Ala	Ser	Pro	Cys
			155						160					165
Arg	Ser	Ser	Asp	Gly	Val	Phe	Tyr	Thr	Gly	Arg	Lys	Gln	Asp	Ala
			170						175					180
Trp	Phe	Val	Val	Asp	Pro	Glu	Ser	Gly	Glu	Thr	Gln	Met	Thr	Leu
			185						190					195
Thr	Thr	Glu	Gly	Pro	Ser	Thr	Pro	Arg	Leu	Tyr	Ile	Gly	Arg	Thr
			200						205					210
Gln	Tyr	Thr	Val	Thr	Met	His	Asp	Pro	Arg	Ala	Pro	Ala	Leu	Arg
			215						220					225
Trp	Asn	Thr	Thr	Tyr	Arg	Arg	Tyr	Ser	Ala	Pro	Pro	Met	Asp	Gly
			230						235					240
Ser	Pro	Gly	Lys	Tyr	Met	Ser	His	Leu	Ala	Ser	Cys	Gly	Met	Gly
			245						250					255
Leu	Leu	Leu	Thr	Val	Asp	Pro	Gly	Ser	Gly	Thr	Val	Leu	Trp	Thr
			260						265					270
Gln	Asp	Leu	Gly	Val	Pro	Val	Met	Gly	Val	Tyr	Thr	Trp	His	Gln
			275						280					285
Asp	Gly	Leu	Arg	Gln	Leu	Pro	His	Leu	Thr	Leu	Ala	Arg	Asp	Thr
			290						295					300
Leu	His	Phe	Leu	Ala	Leu	Arg	Trp	Gly	His	Ile	Arg	Leu	Pro	Ala
			305						310					315
Ser	Gly	Pro	Arg	Asp	Thr	Ala	Thr	Leu	Phe	Ser	Thr	Leu	Asp	Thr
			320						325					330
Gln	Leu	Leu	Met	Thr	Leu	Tyr	Val	Gly	Lys	Asp	Glu	Thr	Gly	Phe
			335						340					345
Tyr	Val	Ser	Lys	Ala	Leu	Val	His	Thr	Gly	Val	Ala	Leu	Val	Pro
			350						355					360
Arg	Gly	Leu	Thr	Leu	Ala	Pro	Ala	Asp	Gly	Pro	Thr	Thr	Asp	Glu
			365						370					375
Val	Thr	Leu	Gln	Val	Ser	Gly	Glu	Arg	Glu	Gly	Ser	Pro	Ser	Thr
			380						385					390
Ala	Val	Arg	Tyr	Pro	Ser	Gly	Ser	Val	Ala	Leu	Pro	Ser	Gln	Trp

	395	400	405
Leu Leu Ile Gly His His Glu Leu Pro	Pro Val Leu His Thr Thr		
	410	415	420
Met Leu Arg Val His Pro Thr Leu Gly	Ser Gly Thr Ala Glu Thr		
	425	430	435
Arg Pro Pro Glu Asn Thr Gln Ala Pro	Ala Phe Phe Leu Glu Leu		
	440	445	450
Leu Ser Leu Ser Arg Glu Lys Leu Trp	Asp Ser Glu Leu His Pro		
	455	460	465
Glu Glu Lys Thr Pro Asp Ser Tyr Leu	Gly Leu Gly Pro Gln Asp		
	470	475	480
Leu Leu Ala Ala Ser Leu Thr Ala Val	Leu Leu Gly Gly Trp Ile		
	485	490	495
Leu Phe Val Met Arg Gln Gln Gln Glu	Thr Pro Leu Ala Pro Ala		
	500	505	510
Asp Phe Ala His Ile Ser Gln Asp Ala	Gln Ser Leu His Ser Gly		
	515	520	525
Ala Ser Arg Arg Ser Gln Lys Arg Leu	Gln Ser Pro Ser Pro Glu		
	530	535	540
Ser Pro Pro Ser Ser Pro Pro Ala Glu	Gln Leu Thr Val Val Gly		
	545	550	555
Lys Ile Ser Phe Asn Pro Lys Asp Val	Leu Gly Arg Gly Ala Gly		
	560	565	570
Gly Thr Phe Val Phe Arg Gly Gln Phe	Glu Gly Arg Ala Val Ala		
	575	580	585
Val Lys Arg Leu Leu Arg Glu Cys Phe	Gly Leu Val Arg Arg Glu		
	590	595	600
Val Gln Leu Leu Gln Glu Ser Asp Arg	His Pro Asn Val Leu Arg		
	605	610	615
Tyr Phe Cys Thr Glu Arg Gly Pro Gln	Phe His Tyr Ile Ala Leu		
	620	625	630
Glu Leu Cys Arg Ala Ser Leu Gln Glu	Tyr Val Glu Asn Pro Asp		
	635	640	645
Leu Asp Arg Gly Gly Leu Glu Pro Glu	Val Val Leu Gln Gln Leu		
	650	655	660
Met Ser Gly Leu Ala His Leu His Ser	Leu His Ile Val His Arg		
	665	670	675
Asp Leu Lys Pro Gly Asn Ile Leu Ile	Thr Gly Pro Asp Ser Gln		
	680	685	690
Gly Leu Gly Arg Val Val Leu Ser Asp	Phe Gly Leu Cys Lys Lys		
	695	700	705
Leu Pro Ala Gly Arg Cys Ser Phe Ser	Leu His Ser Gly Ile Pro		
	710	715	720
Gly Thr Glu Gly Trp Met Ala Pro Glu	Leu Leu Gln Leu Leu Pro		
	725	730	735
Pro Asp Ser Pro Thr Ser Ala Val Asp	Ile Phe Ser Ala Gly Cys		
	740	745	750
Val Phe Tyr Tyr Val Leu Ser Gly Gly	Ser His Pro Phe Gly Asp		
	755	760	765
Ser Leu Tyr Arg Gln Ala Asn Ile Leu	Thr Gly Ala Pro Cys Leu		
	770	775	780
Ala His Leu Glu Glu Glu Val His Asp	Lys Val Val Ala Arg Asp		
	785	790	795
Leu Val Gly Ala Met Leu Ser Pro Leu	Pro Gln Pro Arg Pro Ser		
	800	805	810
Ala Pro Gln Val Leu Ala His Pro Phe	Phe Trp Ser Arg Ala Lys		

	815		820		825
Gln Leu Gln Phe	Phe Gln Asp Val Ser	Asp Trp Leu Glu Lys	Glu		
	830		835		840
Ser Glu Gln Glu	Pro Leu Val Arg Ala	Leu Glu Ala Gly Gly	Cys		
	845		850		855
Ala Val Val Arg	Asp Asn Trp His Glu	His Ile Ser Met Pro	Leu		
	860		865		870
Gln Thr Asp Leu	Arg Lys Phe Arg Ser	Tyr Lys Gly Thr Ser	Val		
	875		880		885
Arg Asp Leu Leu	Arg Ala Val Arg Asn	Lys Lys His His Tyr	Arg		
	890		895		900
Glu Leu Pro Val	Glu Val Arg Gln Ala	Leu Gly Gln Val Pro	Asp		
	905		910		915
Gly Phe Val Gln	Tyr Phe Thr Asn Arg	Phe Pro Arg Leu Leu	Leu		
	920		925		930
His Thr His Arg	Ala Met Arg Ser Cys	Ala Ser Glu Ser Leu	Phe		
	935		940		945
Leu Pro Tyr Tyr	Pro Pro Asp Ser Glu	Ala Arg Arg Pro Cys	Pro		
	950		955		960
Gly Ala Thr Gly	Arg				
	965				

&lt;210&gt; 14

&lt;211&gt; 329

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1851973CD1

&lt;400&gt; 14

Met Asp Pro Thr	Ala Gly Ser Lys Lys	Glu Pro Gly Gly Gly	Ala
1	5	10	15
Ala Thr Glu Glu	Gly Val Asn Arg Ile	Ala Val Pro Lys Pro	Pro
	20	25	30
Ser Ile Glu Glu	Phe Ser Ile Val Lys	Pro Ile Ser Arg Gly	Ala
	35	40	45
Phe Gly Lys Val	Tyr Leu Gly Gln Lys	Gly Gly Lys Leu Tyr	Ala
	50	55	60
Val Lys Val Val	Lys Lys Ala Asp Met	Ile Asn Lys Asn Met	Thr
	65	70	75
His Gln Val Gln	Ala Glu Arg Asp Ala	Leu Ala Leu Ser Lys	Ser
	80	85	90
Pro Phe Ile Val	His Leu Tyr Tyr Ser	Leu Gln Ser Ala Asn	Asn
	95	100	105
Val Tyr Leu Val	Met Glu Tyr Leu Ile	Gly Gly Asp Val Lys	Ser
	110	115	120
Leu Leu His Ile	Tyr Gly Tyr Phe Asp	Glu Glu Met Ala Val	Lys
	125	130	135
Tyr Ile Ser Glu	Val Ala Leu Ala Leu	Asp Tyr Leu His Arg	His
	140	145	150
Gly Ile Ile His	Arg Asp Leu Lys Pro	Asp Asn Met Leu Ile	Ser
	155	160	165
Asn Glu Gly His	Ile Lys Leu Thr Asp	Phe Gly Leu Ser Lys	Val
	170	175	180

Thr	Leu	Asn	Arg	Asp	Ile	Asn	Met	Met	Asp	Ile	Leu	Thr	Thr	Pro
				185					190					195
Ser	Met	Ala	Lys	Pro	Arg	Gln	Asp	Tyr	Ser	Arg	Thr	Pro	Gly	Gln
				200					205					210
Val	Leu	Ser	Leu	Ile	Ser	Ser	Leu	Gly	Phe	Asn	Thr	Pro	Ile	Ala
				215					220					225
Glu	Lys	Asn	Gln	Asp	Pro	Ala	Asn	Ile	Leu	Ser	Ala	Cys	Leu	Ser
				230					235					240
Glu	Thr	Ser	Gln	Leu	Ser	Gln	Gly	Leu	Val	Cys	Pro	Met	Ser	Val
				245					250					255
Asp	Gln	Lys	Asp	Thr	Thr	Pro	Tyr	Ser	Ser	Lys	Leu	Leu	Lys	Ser
				260					265					270
Cys	Leu	Glu	Thr	Val	Ala	Ser	Asn	Pro	Gly	Met	Pro	Val	Lys	Cys
				275					280					285
Leu	Thr	Ser	Asn	Leu	Leu	Gln	Ser	Arg	Lys	Arg	Leu	Ala	Thr	Ser
				290					295					300
Ser	Ala	Ser	Ser	Gln	Ser	His	Thr	Phe	Ile	Ser	Ser	Val	Glu	Ser
				305					310					315
Glu	Cys	His	Ser	Ser	Pro	Lys	Trp	Glu	Lys	Asp	Cys	Gln	Val	
				320					325					

&lt;210&gt; 15

&lt;211&gt; 945

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474604CD1

&lt;400&gt; 15

Met	Thr	Lys	Ser	Glu	Glu	Gln	Gln	Pro	Leu	Ser	Leu	Gln	Lys	Ala
1				5					10					15
Leu	Gln	Gln	Cys	Glu	Leu	Val	Gln	Asn	Met	Ile	Asp	Leu	Ser	Ile
				20					25					30
Ser	Asn	Leu	Glu	Gly	Leu	Arg	Thr	Lys	Cys	Ala	Thr	Ser	Asn	Asp
				35					40					45
Leu	Thr	Gln	Lys	Glu	Ile	Arg	Thr	Leu	Glu	Ser	Lys	Leu	Val	Lys
				50					55					60
Tyr	Phe	Ser	Arg	Gln	Leu	Ser	Cys	Lys	Lys	Lys	Val	Ala	Leu	Gln
				65					70					75
Glu	Arg	Asn	Ala	Glu	Leu	Asp	Gly	Phe	Pro	Gln	Leu	Arg	His	Trp
				80					85					90
Phe	Arg	Ile	Val	Asp	Val	Arg	Lys	Glu	Val	Leu	Glu	Glu	Ile	Ser
				95					100					105
Pro	Gly	Gln	Leu	Ser	Leu	Glu	Asp	Leu	Leu	Glu	Met	Thr	Asp	Glu
				110					115					120
Gln	Val	Cys	Glu	Thr	Val	Glu	Lys	Tyr	Gly	Ala	Asn	Arg	Glu	Glu
				125					130					135
Cys	Ala	Arg	Leu	Asn	Ala	Ser	Leu	Ser	Cys	Leu	Arg	Asn	Val	His
				140					145					150
Met	Ser	Gly	Gly	Asn	Leu	Ser	Lys	Gln	Asp	Trp	Thr	Ile	Gln	Trp
				155					160					165
Pro	Thr	Thr	Glu	Thr	Gly	Lys	Glu	Asn	Asn	Pro	Val	Cys	Pro	Pro
				170					175					180
Glu	Pro	Thr	Pro	Trp	Ile	Arg	Thr	His	Leu	Ser	Gln	Ser	Pro	Arg

	185		190		195
Val Pro Ser Lys Cys Val Gln His Tyr Cys His Thr Ser Pro Thr					
	200		205		210
Pro Gly Ala Pro Val Tyr Thr His Val Asp Arg Leu Thr Val Asp					
	215		220		225
Ala Tyr Pro Gly Leu Cys Pro Pro Pro Pro Leu Glu Ser Gly His					
	230		235		240
Arg Ser Leu Pro Pro Ser Pro Arg Gln Arg His Ala Val Arg Thr					
	245		250		255
Pro Pro Arg Thr Pro Asn Ile Val Thr Thr Val Thr Pro Pro Gly					
	260		265		270
Thr Pro Pro Met Arg Lys Lys Asn Lys Leu Lys Pro Pro Gly Thr					
	275		280		285
Pro Pro Pro Ser Ser Arg Lys Leu Ile His Leu Ile Pro Gly Phe					
	290		295		300
Thr Ala Leu His Arg Ser Lys Ser His Glu Phe Gln Leu Gly His					
	305		310		315
Arg Val Asp Glu Ala His Thr Pro Lys Ala Lys Lys Lys Ser Lys					
	320		325		330
Pro Leu Asn Leu Lys Ile His Ser Ser Val Gly Ser Cys Glu Asn					
	335		340		345
Ile Pro Ser Gln Gln Arg Ser Pro Leu Leu Ser Glu Arg Ser Leu					
	350		355		360
Arg Ser Phe Phe Val Gly His Ala Pro Phe Leu Pro Ser Thr Pro					
	365		370		375
Pro Val His Thr Glu Ala Asn Phe Ser Ala Asn Thr Leu Ser Val					
	380		385		390
Pro Arg Trp Ser Pro Gln Ile Pro Arg Arg Asp Leu Gly Asn Ser					
	395		400		405
Ile Lys His Arg Phe Ser Thr Lys Tyr Trp Met Ser Gln Thr Cys					
	410		415		420
Thr Val Cys Gly Lys Gly Met Leu Phe Gly Leu Lys Cys Lys Asn					
	425		430		435
Cys Lys Leu Lys Cys His Asn Lys Cys Thr Lys Glu Ala Pro Pro					
	440		445		450
Cys His Leu Leu Ile Ile His Arg Gly Asp Pro Ala Arg Leu Val					
	455		460		465
Arg Thr Glu Ser Val Pro Cys Asp Ile Asn Asn Pro Leu Arg Lys					
	470		475		480
Pro Pro Arg Tyr Ser Asp Leu His Ile Ser Gln Thr Leu Pro Lys					
	485		490		495
Thr Asn Lys Ile Asn Lys Asp His Ile Pro Val Pro Tyr Gln Pro					
	500		505		510
Asp Ser Ser Ser Asn Pro Ser Ser Thr Thr Ser Ser Thr Pro Ser					
	515		520		525
Ser Pro Ala Pro Pro Leu Pro Pro Ser Ala Thr Pro Pro Ser Pro					
	530		535		540
Leu His Pro Ser Pro Gln Cys Thr Arg Gln Gln Lys Asn Phe Asn					
	545		550		555
Leu Pro Ala Ser His Tyr Tyr Lys Tyr Lys Gln Gln Phe Ile Phe					
	560		565		570
Pro Asp Val Val Pro Val Pro Glu Thr Pro Thr Arg Ala Pro Gln					
	575		580		585
Val Ile Leu His Pro Val Thr Ser Asn Pro Ile Leu Glu Gly Asn					
	590		595		600
Pro Leu Leu Gln Ile Glu Val Glu Pro Thr Ser Glu Asn Glu Glu					



Val His Asp Glu	605	610	615
Ala Glu Glu Ser Glu		Asp Asp Phe Glu Glu	Met
	620	625	630
Asn Leu Ser Leu		Phe Pro Arg Lys Ala	Ser
Leu Ser Ala Arg Ser	635	640	645
Gln Thr Ser Ile		Asp Ile Pro Phe Glu	Gln
Phe Leu Gln Glu Trp	650	655	660
Leu Glu Ile Gly		Gly Arg Phe Gly Gln	Val
Glu Leu Ile Gly Lys	665	670	675
Tyr His Gly Arg		Ala Ile Arg Leu Ile	Asp
Trp His Gly Glu Val	680	685	690
Ile Glu Arg Asp		Lys Ala Phe Lys Arg	Glu
Asn Glu Asp Gln Leu	695	700	705
Val Met Ala Tyr		Glu Asn Val Val Leu	Phe
Arg Gln Thr Arg His	710	715	720
Met Gly Ala Cys		Leu Ala Ile Ile Thr	Ser
Met Ser Pro Pro His	725	730	735
Leu Cys Lys Gly		Val Val Arg Asp Ala	Lys
Arg Thr Leu Tyr Ser	740	745	750
Ile Val Leu Asp		Gln Ile Ala Gln Glu	Ile
Val Asn Lys Thr Arg	755	760	765
Val Lys Gly Met		Lys Gly Ile Leu His	Lys
Gly Tyr Leu His Ala	770	775	780
Asp Leu Lys Ser		Asp Asn Gly Lys Val	Val
Lys Asn Val Phe Tyr	785	790	795
Ile Thr Asp Phe		Ser Gly Val Leu Gln	Ala
Gly Leu Phe Ser Ile	800	805	810
Gly Arg Arg Glu		Gln Asn Gly Trp Leu	Cys
Asp Lys Leu Arg Ile	815	820	825
His Leu Ala Pro		Leu Ser Pro Asp Thr	Glu
Glu Ile Ile Arg Gln	830	835	840
Glu Asp Lys Leu		Ser Asp Val Phe Ala	Leu
Pro Phe Ser Lys His	845	850	855
Gly Thr Ile Trp		Arg Glu Trp Pro Phe	Lys
Tyr Glu Leu His Ala	860	865	870
Thr Gln Pro Ala		Gln Met Gly Thr Gly	Met
Glu Ala Ile Ile Trp	875	880	885
Lys Pro Asn Leu		Gly Lys Glu Ile Ser	Asp
Ser Gln Ile Gly Met	890	895	900
Ile Leu Leu Phe		Gln Glu Glu Arg Pro	Thr
Cys Trp Ala Phe Glu	905	910	915
Phe Thr Lys Leu		Lys Leu Pro Lys Arg	Asn
Met Asp Met Leu Glu	920	925	930
Arg Arg Leu Ser		Trp Lys Ser Ala Glu	Leu
His Pro Gly His Phe	935	940	945

&lt;210&gt; 16

&lt;211&gt; 1009

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474721CD1

&lt;400&gt; 16

```

Met Glu Thr Cys Ala Gly Pro His Pro Leu Arg Leu Phe Leu Cys
 1              5              10              15
Arg Met Gln Leu Cys Leu Ala Leu Leu Leu Gly Pro Trp Arg Pro
      20              25              30
Gly Thr Ala Glu Glu Val Ile Leu Leu Asp Ser Lys Ala Ser Gln
      35              40              45
Ala Glu Leu Gly Trp Thr Ala Leu Pro Ser Asn Gly Trp Glu Glu
      50              55              60
Ile Ser Gly Val Asp Glu His Asp Arg Pro Ile Arg Thr Tyr Gln
      65              70              75
Val Cys Asn Val Leu Glu Pro Asn Gln Asp Asn Trp Leu Gln Thr
      80              85              90
Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg Ile Phe Val Glu Leu
      95              100             105
Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro Gly Ala Ala Gly
      110             115             120
Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu Thr Glu Ala
      125             130             135
Asp Leu Gly Arg Gly Arg Pro Arg Leu Gly Gly Ser Arg Pro Arg
      140             145             150
Lys Ile Asp Thr Ile Ala Ala Asp Glu Ser Phe Thr Gln Gly Asp
      155             160             165
Leu Gly Glu Arg Lys Met Lys Leu Asn Thr Glu Val Arg Glu Ile
      170             175             180
Gly Pro Leu Ser Arg Arg Gly Phe His Leu Ala Phe Gln Asp Val
      185             190             195
Gly Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Tyr Lys Gln
      200             205             210
Cys Arg Ala Thr Val Arg Gly Leu Ala Thr Phe Pro Ala Thr Ala
      215             220             225
Ala Glu Ser Ala Phe Ser Thr Leu Val Glu Val Ala Gly Thr Cys
      230             235             240
Val Ala His Ser Glu Gly Glu Pro Gly Ser Pro Pro Arg Met His
      245             250             255
Cys Gly Ala Asp Gly Glu Trp Leu Val Pro Val Gly Arg Cys Ser
      260             265             270
Cys Ser Ala Gly Phe Gln Glu Arg Gly Asp Ile Cys Glu Ala Cys
      275             280             285
Pro Pro Gly Phe Tyr Lys Val Ser Pro Arg Arg Arg Val Cys Ser
      290             295             300
Pro Cys Pro Glu His Ser Arg Ala Leu Glu Asn Ala Ser Thr Phe
      305             310             315
Cys Val Cys Gln Asp Ser Tyr Ala Arg Ser Pro Thr Asp Pro Pro
      320             325             330
Ser Ala Ser Cys Thr Arg Gly Pro Pro Ser Ala Pro Arg Asp Leu
      335             340             345
Gln Tyr Ser Leu Ser Arg Ser Pro Leu Val Leu Arg Leu Arg Trp
      350             355             360
Leu Pro Pro Ala Asp Ser Gly Gly Arg Ser Asp Val Thr Tyr Ser
      365             370             375
Leu Leu Cys Leu Arg Cys Gly Arg Glu Gly Pro Ala Gly Ala Cys
      380             385             390
Glu Pro Cys Gly Pro Arg Val Ala Phe Leu Pro Arg Gln Ala Gly
      395             400             405
Leu Arg Glu Arg Ala Ala Thr Leu Leu His Leu Arg Pro Gly Ala

```

Arg Tyr Thr Val	410	415	420
Arg Val Ala Val Leu Asn Gly Val Ser Gly Pro	425	430	435
Ala Ala Ala Leu Val Pro Val Gly Ala Val Ser Ile Asn Pro Gly	440	445	450
Thr Val Gly Pro Val Pro Val Ala Gly Val Ile Arg Asp Arg Val	455	460	465
Glu Pro Gln Ser Val Ser Leu Ser Trp Arg Glu Pro Ile Pro Ala	470	475	480
Gly Ala Pro Gly Ala Asn Asp Thr Glu Tyr Glu Ile Arg Tyr Tyr	485	490	495
Glu Lys Val Gln Ser Glu Gln Thr Tyr Ser Met Val Lys Thr Gly	500	505	510
Ala Pro Thr Val Thr Val Thr Asn Leu Lys Pro Ala Thr Arg Tyr	515	520	525
Val Phe Gln Ile Arg Ala Ala Ser Pro Gly Pro Ser Trp Glu Ala	530	535	540
Gln Ser Phe Asn Pro Ser Ile Glu Val Gln Thr Leu Gly Glu Ala	545	550	555
Ala Ser Gly Ser Arg Asp Gln Ser Pro Ala Ile Val Val Thr Val	560	565	570
Val Thr Ile Ser Ala Leu Leu Val Leu Gly Ser Val Met Ser Val	575	580	585
Leu Ala Ile Trp Arg Arg Pro Cys Ser Tyr Gly Lys Gly Gly Gly	590	595	600
Asp Ala His Asp Glu Glu Glu Leu Tyr Phe His Phe Lys Val Pro	605	610	615
Thr Arg Arg Thr Phe Leu Asp Pro Gln Ser Cys Gly Asp Leu Leu	620	625	630
Gln Ala Val His Leu Phe Ala Lys Glu Leu Asp Ala Lys Ser Val	635	640	645
Thr Leu Glu Arg Ser Leu Gly Gly Gly Arg Phe Gly Glu Leu Cys	650	655	660
Cys Gly Cys Leu Gln Leu Pro Gly Arg Gln Glu Leu Leu Val Ala	665	670	675
Val His Met Leu Arg Asp Ser Ala Ser Asp Ser Gln Arg Leu Gly	680	685	690
Phe Leu Ala Glu Ala Leu Thr Leu Gly Gln Phe Asp His Ser His	695	700	705
Ile Val Arg Leu Glu Gly Val Val Thr Arg Gly Ser Thr Leu Met	710	715	720
Ile Val Thr Glu Tyr Met Ser His Gly Ala Leu Asp Gly Phe Leu	725	730	735
Arg Arg His Glu Gly Gln Leu Val Ala Gly Gln Leu Met Gly Leu	740	745	750
Leu Pro Gly Leu Ala Ser Ala Met Lys Tyr Leu Ser Glu Met Gly	755	760	765
Tyr Val His Arg Gly Leu Ala Ala Arg His Val Leu Val Ser Ser	770	775	780
Asp Leu Val Cys Lys Ile Ser Gly Phe Gly Arg Gly Pro Arg Asp	785	790	795
Arg Ser Glu Ala Val Tyr Thr Thr Met Ser Gly Arg Ser Pro Ala	800	805	810
Leu Trp Ala Ala Pro Glu Thr Leu Gln Phe Gly His Phe Ser Ser	815	820	825
Ala Ser Asp Val Trp Ser Phe Gly Ile Ile Met Trp Glu Val Met			

830	835	840
Ala Phe Gly Glu Arg Pro Tyr Trp Asp Met Ser Gly Gln Asp Val		
845	850	855
Ile Lys Ala Val Glu Asp Gly Phe Arg Leu Pro Pro Pro Arg Asn		
860	865	870
Cys Pro Asn Leu Leu His Arg Leu Met Leu Asp Cys Trp Gln Lys		
875	880	885
Asp Pro Gly Glu Arg Pro Arg Phe Ser Gln Ile His Ser Ile Leu		
890	895	900
Ser Lys Met Val Gln Asp Pro Glu Pro Pro Lys Cys Ala Leu Thr		
905	910	915
Thr Cys Pro Arg Pro Pro Thr Pro Leu Ala Asp Arg Ala Phe Ser		
920	925	930
Thr Phe Pro Ser Phe Gly Ser Val Gly Ala Trp Leu Glu Ala Leu		
935	940	945
Asp Leu Cys Arg Tyr Lys Asp Ser Phe Ala Ala Ala Gly Tyr Gly		
950	955	960
Ser Leu Glu Ala Val Ala Glu Met Thr Ala Gln Arg Asp Leu Val		
965	970	975
Ser Leu Gly Ile Ser Leu Ala Glu His Arg Glu Ala Leu Leu Ser		
980	985	990
Gly Ile Ser Ala Leu Gln Ala Arg Val Leu Gln Leu Gln Gly Gln		
995	1000	1005
Gly Val Gln Val		

&lt;210&gt; 17

&lt;211&gt; 917

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478815CD1

&lt;400&gt; 17

Met Phe Ala Val His Leu Met Ala Phe Tyr Phe Ser Lys Leu Lys		
1	5	10
Glu Asp Gln Ile Lys Lys Val Asp Arg Phe Leu Tyr His Met Arg		
20	25	30
Leu Ser Asp Asp Thr Leu Leu Asp Ile Met Arg Arg Phe Arg Ala		
35	40	45
Glu Met Glu Lys Gly Leu Ala Lys Asp Thr Asn Pro Thr Ala Ala		
50	55	60
Val Lys Met Leu Pro Thr Phe Val Arg Ala Ile Pro Asp Gly Ser		
65	70	75
Glu Asn Gly Glu Phe Leu Ser Leu Asp Leu Gly Gly Ser Lys Phe		
80	85	90
Arg Val Leu Lys Val Gln Val Ala Glu Glu Gly Lys Arg His Val		
95	100	105
Gln Met Glu Ser Gln Phe Tyr Pro Thr Pro Asn Glu Ile Ile Arg		
110	115	120
Gly Asn Gly Thr Glu Leu Phe Glu Tyr Val Ala Asp Cys Leu Ala		
125	130	135
Asp Phe Met Lys Thr Lys Asp Leu Lys His Lys Lys Leu Pro Leu		
140	145	150

Gly Leu Thr Phe Ser Phe Pro Cys Arg Gln Thr Lys Leu Glu Glu	155	160	165
Gly Val Leu Leu Ser Trp Thr Lys Lys Phe Lys Ala Arg Gly Val	170	175	180
Gln Asp Thr Asp Val Val Ser Arg Leu Thr Lys Ala Met Arg Arg	185	190	195
His Lys Asp Met Asp Val Asp Ile Leu Ala Leu Val Asn Asp Thr	200	205	210
Val Gly Thr Met Met Thr Cys Ala Tyr Asp Asp Pro Tyr Cys Glu	215	220	225
Val Gly Val Ile Ile Gly Thr Gly Thr Asn Ala Cys Tyr Met Glu	230	235	240
Asp Met Ser Asn Ile Asp Leu Val Glu Gly Asp Glu Gly Arg Met	245	250	255
Cys Ile Asn Thr Glu Trp Gly Ala Phe Gly Asp Asp Gly Ala Leu	260	265	270
Glu Asp Ile Arg Thr Glu Phe Asp Arg Glu Leu Asp Leu Gly Ser	275	280	285
Leu Asn Pro Gly Lys Gln Leu Phe Glu Lys Met Ile Ser Gly Leu	290	295	300
Tyr Leu Gly Glu Leu Val Arg Leu Ile Leu Leu Lys Met Ala Lys	305	310	315
Ala Gly Leu Leu Phe Gly Gly Glu Lys Ser Ser Ala Leu His Thr	320	325	330
Lys Gly Lys Ile Glu Thr Arg His Val Ala Ala Met Glu Lys Tyr	335	340	345
Lys Glu Gly Leu Ala Asn Thr Arg Glu Ile Leu Val Asp Leu Gly	350	355	360
Leu Glu Pro Ser Glu Ala Asp Cys Ile Ala Val Gln His Val Cys	365	370	375
Thr Ile Val Ser Phe Arg Ser Ala Asn Leu Cys Ala Ala Ala Leu	380	385	390
Ala Ala Ile Leu Thr Arg Leu Arg Glu Asn Lys Lys Val Glu Arg	395	400	405
Leu Arg Thr Thr Val Gly Met Asp Gly Thr Leu Tyr Lys Ile His	410	415	420
Pro Gln Tyr Pro Lys Arg Leu His Lys Val Val Arg Lys Leu Val	425	430	435
Pro Ser Cys Asp Val Arg Phe Leu Leu Ser Glu Ser Gly Ser Thr	440	445	450
Lys Gly Ala Ala Met Val Thr Ala Val Ala Ser Arg Val Gln Ala	455	460	465
Gln Arg Lys Gln Ile Asp Arg Val Leu Ala Leu Phe Gln Leu Thr	470	475	480
Arg Glu Gln Leu Val Asp Val Gln Ala Lys Met Arg Ala Glu Leu	485	490	495
Glu Tyr Gly Leu Lys Lys Lys Ser His Gly Leu Ala Thr Val Arg	500	505	510
Met Leu Pro Thr Tyr Val Cys Gly Leu Pro Asp Gly Thr Glu Lys	515	520	525
Gly Lys Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg Val	530	535	540
Leu Leu Val Lys Ile Arg Ser Gly Arg Arg Ser Val Arg Met Tyr	545	550	555
Asn Lys Ile Phe Ala Ile Pro Leu Glu Ile Met Gln Gly Thr Gly	560	565	570

Glu	Glu	Leu	Phe	Asp	His	Ile	Val	Gln	Cys	Ile	Ala	Asp	Phe	Leu	575	580	585
Asp	Tyr	Met	Gly	Leu	Lys	Gly	Ala	Ser	Leu	Pro	Leu	Gly	Phe	Thr	590	595	600
Phe	Ser	Phe	Pro	Cys	Arg	Gln	Met	Ser	Ile	Asp	Lys	Gly	Thr	Leu	605	610	615
Ile	Gly	Trp	Thr	Lys	Gly	Phe	Lys	Ala	Thr	Asp	Cys	Glu	Gly	Glu	620	625	630
Asp	Val	Val	Asp	Met	Leu	Arg	Glu	Ala	Ile	Lys	Arg	Arg	Asn	Glu	635	640	645
Phe	Asp	Leu	Asp	Ile	Val	Ala	Val	Val	Asn	Asp	Thr	Val	Gly	Thr	650	655	660
Met	Met	Thr	Cys	Gly	Tyr	Glu	Asp	Pro	Asn	Cys	Glu	Ile	Gly	Leu	665	670	675
Ile	Ala	Gly	Thr	Gly	Ser	Asn	Met	Cys	Tyr	Met	Glu	Asp	Met	Arg	680	685	690
Asn	Ile	Glu	Met	Val	Glu	Gly	Gly	Glu	Gly	Lys	Met	Cys	Ile	Asn	695	700	705
Thr	Glu	Trp	Gly	Gly	Phe	Gly	Asp	Asn	Gly	Cys	Ile	Asp	Asp	Ile	710	715	720
Arg	Thr	Arg	Tyr	Asp	Thr	Glu	Val	Asp	Glu	Gly	Ser	Leu	Asn	Pro	725	730	735
Gly	Lys	Gln	Arg	Tyr	Glu	Lys	Met	Thr	Ser	Gly	Met	Tyr	Leu	Gly	740	745	750
Glu	Ile	Val	Arg	Gln	Ile	Leu	Ile	Asp	Leu	Thr	Lys	Gln	Gly	Leu	755	760	765
Leu	Phe	Arg	Gly	Gln	Ile	Ser	Glu	Arg	Leu	Arg	Thr	Arg	Gly	Ile	770	775	780
Phe	Glu	Thr	Lys	Phe	Leu	Ser	Gln	Ile	Glu	Ser	Asp	Arg	Leu	Ala	785	790	795
Leu	Leu	Gln	Val	Arg	Arg	Ile	Leu	Gln	Gln	Leu	Gly	Leu	Asp	Ser	800	805	810
Thr	Cys	Glu	Asp	Ser	Ile	Val	Val	Lys	Glu	Val	Cys	Gly	Ala	Val	815	820	825
Ser	Arg	Arg	Ala	Ala	Gln	Leu	Cys	Gly	Ala	Gly	Leu	Ala	Ala	Ile	830	835	840
Val	Glu	Lys	Arg	Arg	Glu	Asp	Gln	Gly	Leu	Glu	His	Leu	Arg	Ile	845	850	855
Thr	Val	Gly	Val	Asp	Gly	Thr	Leu	Tyr	Lys	Leu	His	Pro	His	Phe	860	865	870
Ser	Arg	Ile	Leu	Gln	Glu	Thr	Val	Lys	Glu	Leu	Ala	Pro	Arg	Cys	875	880	885
Asp	Val	Thr	Phe	Met	Leu	Ser	Glu	Asp	Gly	Ser	Gly	Lys	Gly	Ala	890	895	900
Ala	Leu	Ile	Thr	Ala	Val	Ala	Lys	Arg	Leu	Gln	Gln	Ala	Gln	Lys	905	910	915

Glu Asn

&lt;210&gt; 18

&lt;211&gt; 2380

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477141CD1

&lt;400&gt; 18

Met	Asn	His	Pro	Pro	Trp	Pro	Ser	Leu	Asp	Cys	His	Leu	Lys	Ala	
1				5					10					15	
Arg	Ser	Gly	His	Ala	Leu	Leu	Ser	Trp	Pro	Gly	Gly	Trp	Ala	Phe	
				20					25					30	
Pro	Ile	Ser	Arg	Glu	Gln	Asn	Ala	Ser	Leu	Ser	Leu	Cys	Leu	Ser	
				35					40					45	
Val	Ser	Leu	Cys	Val	Arg	Met	Cys	Val	Ser	Leu	Thr	Leu	Cys	Val	
				50					55					60	
Ser	Ala	Leu	Cys	Val	Ala	Pro	Val	Ala	Ala	Phe	Pro	Ser	Ala	His	
				65					70					75	
Pro	Glu	Ser	Arg	Ser	Leu	Ala	Val	Leu	Ala	Pro	Leu	Gln	Asp	Val	
				80					85					90	
Asp	Val	Gly	Ala	Gly	Glu	Met	Ala	Leu	Phe	Glu	Cys	Leu	Val	Ala	
				95					100					105	
Gly	Pro	Thr	Asp	Val	Glu	Val	Asp	Trp	Leu	Cys	Arg	Gly	Arg	Leu	
				110					115					120	
Leu	Gln	Pro	Ala	Leu	Leu	Lys	Cys	Lys	Met	His	Phe	Asp	Gly	Arg	
				125					130					135	
Lys	Cys	Lys	Leu	Leu	Leu	Thr	Ser	Val	His	Glu	Asp	Asp	Ser	Gly	
				140					145					150	
Val	Tyr	Thr	Cys	Lys	Leu	Ser	Thr	Ala	Lys	Asp	Glu	Leu	Thr	Cys	
				155					160					165	
Ser	Ala	Arg	Leu	Thr	Val	Arg	Pro	Ser	Leu	Ala	Pro	Leu	Phe	Thr	
				170					175					180	
Arg	Leu	Leu	Glu	Asp	Val	Glu	Val	Leu	Glu	Gly	Arg	Ala	Ala	Arg	
				185					190					195	
Phe	Asp	Cys	Lys	Ile	Ser	Gly	Thr	Pro	Pro	Pro	Val	Val	Thr	Trp	
				200					205					210	
Thr	His	Phe	Gly	Cys	Pro	Met	Glu	Glu	Ser	Glu	Asn	Leu	Arg	Leu	
				215					220					225	
Arg	Gln	Asp	Gly	Gly	Leu	His	Ser	Leu	His	Ile	Ala	His	Val	Gly	
				230					235					240	
Ser	Glu	Asp	Glu	Gly	Leu	Tyr	Ala	Val	Ser	Ala	Val	Asn	Thr	His	
				245					250					255	
Gly	Gln	Ala	His	Cys	Ser	Ala	Gln	Leu	Tyr	Val	Glu	Glu	Pro	Arg	
				260					265					270	
Thr	Ala	Ala	Ser	Gly	Pro	Ser	Ser	Lys	Leu	Glu	Lys	Met	Pro	Ser	
				275					280					285	
Ile	Pro	Glu	Glu	Pro	Glu	Gln	Gly	Glu	Leu	Glu	Arg	Leu	Ser	Ile	
				290					295					300	
Pro	Asp	Phe	Leu	Arg	Pro	Leu	Gln	Asp	Leu	Glu	Val	Gly	Leu	Ala	
				305					310					315	
Lys	Glu	Ala	Met	Leu	Glu	Cys	Gln	Val	Thr	Gly	Leu	Pro	Tyr	Pro	
				320					325					330	
Thr	Ile	Ser	Trp	Phe	His	Asn	Gly	His	Arg	Ile	Gln	Ser	Ser	Asp	
				335					340					345	
Asp	Arg	Arg	Met	Thr	Gln	Tyr	Arg	Asp	Val	His	Arg	Leu	Val	Phe	
				350					355					360	
Pro	Ala	Val	Gly	Pro	Gln	His	Ala	Gly	Val	Tyr	Lys	Ser	Val	Ile	
				365					370					375	
Ala	Asn	Lys	Leu	Gly	Lys	Ala	Ala	Cys	Tyr	Ala	His	Leu	Tyr	Val	
				380					385					390	
Thr	Asp	Val	Val	Pro	Gly	Pro	Pro	Asp	Gly	Ala	Pro	Gln	Val	Val	

	395		400		405
Ala Val Thr Gly Arg Met Val Thr Leu Thr Trp Asn Pro Pro Arg					
	410		415		420
Ser Leu Asp Met Ala Ile Asp Pro Asp Ser Leu Thr Tyr Thr Val					
	425		430		435
Gln His Gln Val Leu Gly Ser Asp Gln Trp Thr Ala Leu Val Thr					
	440		445		450
Gly Leu Arg Glu Pro Gly Trp Ala Ala Thr Gly Leu Arg Lys Gly					
	455		460		465
Val Gln His Ile Phe Arg Val Leu Ser Thr Thr Val Lys Ser Ser					
	470		475		480
Ser Lys Pro Ser Pro Pro Ser Glu Pro Val Gln Leu Leu Glu His					
	485		490		495
Gly Pro Thr Leu Glu Glu Ala Pro Ala Met Leu Asp Lys Pro Asp					
	500		505		510
Ile Val Tyr Val Val Glu Gly Gln Pro Ala Ser Val Thr Val Thr					
	515		520		525
Phe Asn His Val Glu Ala Gln Val Val Trp Arg Ser Cys Arg Gly					
	530		535		540
Ala Leu Leu Glu Ala Arg Ala Gly Val Tyr Glu Leu Ser Gln Pro					
	545		550		555
Asp Asp Asp Gln Tyr Cys Leu Arg Ile Cys Arg Val Ser Arg Arg					
	560		565		570
Asp Met Gly Ala Leu Thr Cys Thr Ala Arg Asn Arg His Gly Thr					
	575		580		585
Gln Thr Cys Ser Val Thr Leu Glu Leu Ala Glu Ala Pro Arg Phe					
	590		595		600
Glu Ser Ile Met Glu Asp Val Glu Val Gly Ala Gly Glu Thr Ala					
	605		610		615
Arg Phe Ala Val Val Val Glu Gly Lys Pro Leu Pro Asp Ile Met					
	620		625		630
Trp Tyr Lys Asp Glu Val Leu Leu Thr Glu Ser Ser His Val Ser					
	635		640		645
Phe Val Tyr Glu Glu Asn Glu Cys Ser Leu Val Val Leu Ser Thr					
	650		655		660
Gly Ala Gln Asp Gly Gly Val Tyr Thr Cys Thr Ala Gln Asn Leu					
	665		670		675
Ala Gly Glu Val Ser Cys Lys Ala Glu Leu Ala Val His Ser Ala					
	680		685		690
Gln Thr Ala Met Glu Val Glu Gly Val Gly Glu Asp Glu Asp His					
	695		700		705
Arg Gly Arg Arg Leu Ser Asp Phe Tyr Asp Ile His Gln Glu Ile					
	710		715		720
Gly Arg Gly Ala Phe Ser Tyr Leu Arg Arg Ile Val Glu Arg Ser					
	725		730		735
Ser Gly Leu Glu Phe Ala Ala Lys Phe Ile Pro Ser Gln Ala Lys					
	740		745		750
Pro Lys Ala Ser Ala Arg Arg Glu Ala Arg Leu Leu Ala Arg Leu					
	755		760		765
Gln His Asp Cys Val Leu Tyr Phe His Glu Ala Phe Glu Arg Arg					
	770		775		780
Arg Gly Leu Val Ile Val Thr Glu Leu Cys Thr Glu Glu Leu Leu					
	785		790		795
Glu Arg Ile Ala Arg Lys Pro Thr Val Cys Glu Ser Glu Ile Arg					
	800		805		810
Ala Tyr Met Arg Gln Val Leu Glu Gly Ile His Tyr Leu His Gln					



	815		820		825
Ser His Val Leu His	Leu Asp Val Lys	Pro Glu Asn Leu Leu Val			
	830		835		840
Trp Asp Gly Ala Ala	Gly Glu Gln Gln Val Arg Ile Cys Asp Phe				
	845		850		855
Gly Asn Ala Gln Glu	Leu Thr Pro Gly Glu Pro Gln Tyr Cys Gln				
	860		865		870
Tyr Gly Thr Pro Glu	Phe Val Ala Pro Glu Ile Val Asn Gln Ser				
	875		880		885
Pro Val Ser Gly Val	Thr Asp Ile Trp Pro Val Gly Val Val Ala				
	890		895		900
Phe Leu Cys Leu Thr	Gly Ile Ser Pro Phe Val Gly Glu Asn Asp				
	905		910		915
Arg Thr Thr Leu Met	Asn Ile Arg Asn Tyr Asn Val Ala Phe Glu				
	920		925		930
Glu Thr Thr Phe Leu	Ser Leu Ser Arg Glu Ala Arg Gly Phe Leu				
	935		940		945
Ile Lys Val Leu Val	Gln Asp Arg Leu Arg Pro Thr Ala Glu Glu				
	950		955		960
Thr Leu Glu His Pro	Trp Phe Lys Thr Gln Ala Lys Gly Ala Glu				
	965		970		975
Val Ser Thr Asp His	Leu Lys Leu Phe Leu Ser Arg Arg Arg Trp				
	980		985		990
Gln Arg Ser Gln Ile	Ser Tyr Lys Cys His Leu Val Leu Arg Pro				
	995		1000		1005
Ile Pro Glu Leu Leu	Arg Ala Pro Pro Glu Arg Val Trp Val Thr				
	1010		1015		1020
Met Pro Arg Arg Pro	Pro Pro Ser Gly Gly Leu Ser Ser Ser Ser				
	1025		1030		1035
Asp Ser Glu Glu Glu	Glu Leu Glu Glu Leu Pro Ser Val Pro Arg				
	1040		1045		1050
Pro Leu Gln Pro Glu	Phe Ser Gly Ser Arg Val Ser Leu Thr Asp				
	1055		1060		1065
Ile Pro Thr Glu Asp	Glu Ala Leu Gly Thr Pro Glu Thr Gly Ala				
	1070		1075		1080
Ala Thr Pro Met Asp	Trp Gln Glu Gln Gly Arg Ala Pro Ser Gln				
	1085		1090		1095
Asp Gln Glu Ala Pro	Ser Pro Glu Ala Leu Pro Ser Pro Gly Gln				
	1100		1105		1110
Glu Pro Ala Ala Gly	Ala Ser Pro Arg Arg Gly Glu Leu Arg Arg				
	1115		1120		1125
Gly Ser Ser Ala Glu	Ser Ala Leu Pro Arg Ala Gly Pro Arg Glu				
	1130		1135		1140
Leu Gly Arg Gly Leu	His Lys Ala Ala Ser Val Glu Leu Pro Gln				
	1145		1150		1155
Arg Arg Ser Pro Gly	Pro Gly Ala Thr Arg Leu Ala Arg Gly Gly				
	1160		1165		1170
Leu Gly Glu Gly Glu	Tyr Ala Gln Arg Leu Gln Ala Leu Arg Gln				
	1175		1180		1185
Arg Leu Leu Arg Gly	Gly Pro Glu Asp Gly Lys Val Ser Gly Leu				
	1190		1195		1200
Arg Gly Pro Leu Leu	Glu Ser Leu Gly Gly Arg Ala Arg Asp Pro				
	1205		1210		1215
Arg Met Ala Arg Ala	Ser Ser Glu Ala Ala Pro His His Gln				
	1220		1225		1230
Pro Pro Leu Glu Asn	Arg Gly Leu Gln Lys Ser Ser Ser Phe Ser				

1235	1240	1245
Gln Gly Glu Ala Glu Pro Arg Gly Arg His Arg Arg Ala Gly Ala		
1250	1255	1260
Pro Leu Glu Ile Pro Val Ala Arg Leu Gly Ala Arg Arg Leu Gln		
1265	1270	1275
Glu Ser Pro Ser Leu Ser Ala Leu Ser Glu Ala Gln Pro Ser Ser		
1280	1285	1290
Pro Ala Arg Pro Ser Ala Pro Lys Pro Ser Thr Pro Lys Ser Ala		
1295	1300	1305
Glu Pro Ser Ala Thr Thr Pro Ser Asp Ala Pro Gln Pro Pro Ala		
1310	1315	1320
Pro Gln Pro Ala Gln Asp Lys Ala Pro Glu Pro Arg Pro Glu Pro		
1325	1330	1335
Val Arg Ala Ser Lys Pro Ala Pro Pro Pro Gln Ala Leu Gln Thr		
1340	1345	1350
Leu Ala Leu Pro Leu Thr Pro Tyr Ala Gln Ile Ile Gln Ser Leu		
1355	1360	1365
Gln Leu Ser Gly His Ala Gln Gly Pro Ser Gln Gly Pro Ala Ala		
1370	1375	1380
Pro Pro Ser Glu Pro Lys Pro His Ala Ala Val Phe Ala Arg Val		
1385	1390	1395
Ala Ser Pro Pro Pro Gly Ala Pro Glu Lys Arg Val Pro Ser Ala		
1400	1405	1410
Gly Gly Pro Pro Val Leu Ala Glu Lys Ala Arg Val Pro Thr Val		
1415	1420	1425
Pro Pro Arg Pro Gly Ser Ser Leu Ser Ser Ser Ile Glu Asn Leu		
1430	1435	1440
Glu Ser Glu Ala Val Phe Glu Ala Lys Phe Lys Arg Ser Arg Glu		
1445	1450	1455
Ser Pro Leu Ser Leu Gly Leu Arg Leu Leu Ser Arg Ser Arg Ser		
1460	1465	1470
Glu Glu Arg Gly Pro Phe Arg Gly Ala Glu Glu Glu Asp Gly Ile		
1475	1480	1485
Tyr Arg Pro Ser Pro Ala Gly Thr Pro Leu Glu Leu Val Arg Arg		
1490	1495	1500
Pro Glu Arg Ser Arg Ser Val Gln Asp Leu Arg Ala Val Gly Glu		
1505	1510	1515
Pro Gly Leu Val Arg Arg Leu Ser Leu Ser Leu Ser Gln Arg Leu		
1520	1525	1530
Arg Arg Thr Pro Pro Ala Gln Arg His Pro Ala Trp Glu Ala Arg		
1535	1540	1545
Gly Gly Asp Gly Glu Ser Ser Glu Gly Ser Ser Ala Arg Gly		
1550	1555	1560
Ser Pro Val Leu Ala Met Arg Arg Arg Leu Ser Phe Thr Leu Glu		
1565	1570	1575
Arg Leu Ser Ser Arg Leu Gln Arg Ser Gly Ser Ser Glu Asp Ser		
1580	1585	1590
Gly Gly Ala Ser Gly Arg Ser Thr Pro Leu Phe Gly Arg Leu Arg		
1595	1600	1605
Arg Ala Thr Ser Glu Gly Glu Ser Leu Arg Arg Leu Gly Leu Pro		
1610	1615	1620
His Asn Gln Leu Ala Ala Gln Ala Gly Ala Thr Thr Pro Ser Ala		
1625	1630	1635
Glu Ser Leu Gly Ser Glu Ala Ser Ala Thr Ser Gly Ser Ser Ala		
1640	1645	1650
Pro Gly Glu Ser Arg Ser Arg Leu Arg Trp Gly Phe Ser Arg Pro		

1655	1660	1665
Arg Lys Asp Lys Gly Leu Ser Pro Pro Asn Leu Ser Ala Ser Val		
1670	1675	1680
Gln Glu Glu Leu Gly His Gln Tyr Val Arg Ser Glu Ser Asp Phe		
1685	1690	1695
Pro Pro Val Phe His Ile Lys Leu Lys Asp Gln Val Leu Leu Glu		
1700	1705	1710
Gly Glu Ala Ala Thr Leu Leu Cys Leu Pro Ala Ala Cys Pro Ala		
1715	1720	1725
Pro His Ile Ser Trp Met Lys Asp Lys Lys Ser Leu Arg Ser Glu		
1730	1735	1740
Pro Ser Val Ile Ile Val Ser Cys Lys Asp Gly Arg Gln Leu Leu		
1745	1750	1755
Ser Ile Pro Arg Ala Gly Lys Arg His Ala Gly Leu Tyr Glu Cys		
1760	1765	1770
Ser Ala Thr Asn Val Leu Gly Ser Ile Thr Ser Ser Cys Thr Val		
1775	1780	1785
Ala Val Ala Arg Val Pro Gly Lys Leu Ala Pro Pro Glu Val Pro		
1790	1795	1800
Gln Thr Tyr Gln Asp Thr Ala Leu Val Leu Trp Lys Pro Gly Asp		
1805	1810	1815
Ser Arg Ala Pro Cys Thr Tyr Thr Leu Glu Arg Arg Val Asp Gly		
1820	1825	1830
Glu Ser Val Trp His Pro Val Ser Ser Gly Ile Pro Asp Cys Tyr		
1835	1840	1845
Tyr Asn Val Thr His Leu Pro Val Gly Val Thr Val Arg Phe Arg		
1850	1855	1860
Val Ala Cys Ala Asn Arg Ala Gly Gln Gly Pro Phe Ser Asn Ser		
1865	1870	1875
Ser Glu Lys Val Phe Val Arg Gly Thr Gln Asp Ser Ser Ala Val		
1880	1885	1890
Pro Ser Ala Ala His Gln Glu Ala Pro Val Thr Ser Arg Pro Ala		
1895	1900	1905
Arg Ala Arg Pro Pro Asp Ser Pro Thr Ser Leu Ala Pro Pro Leu		
1910	1915	1920
Ala Pro Ala Ala Pro Thr Pro Pro Ser Val Thr Val Ser Pro Ser		
1925	1930	1935
Ser Pro Pro Thr Pro Pro Ser Gln Ala Leu Ser Ser Leu Lys Ala		
1940	1945	1950
Val Gly Pro Pro Pro Gln Thr Pro Pro Arg Arg His Arg Gly Leu		
1955	1960	1965
Gln Ala Ala Arg Pro Ala Glu Pro Thr Leu Pro Ser Thr His Val		
1970	1975	1980
Thr Pro Ser Glu Pro Lys Pro Phe Val Leu Asp Thr Gly Thr Pro		
1985	1990	1995
Ile Pro Ala Ser Thr Pro Gln Gly Val Lys Pro Val Ser Ser Ser		
2000	2005	2010
Thr Pro Val Tyr Val Val Thr Ser Phe Val Ser Ala Pro Pro Ala		
2015	2020	2025
Pro Glu Pro Pro Ala Pro Glu Pro Pro Pro Glu Pro Thr Lys Val		
2030	2035	2040
Thr Val Gln Ser Leu Ser Pro Ala Lys Glu Val Val Ser Ser Pro		
2045	2050	2055
Gly Ser Ser Pro Arg Ser Ser Pro Arg Pro Glu Gly Thr Thr Leu		
2060	2065	2070
Arg Gln Gly Pro Pro Gln Lys Pro Tyr Thr Phe Leu Glu Glu Lys		

2075	2080	2085
Ala Arg Gly Arg Phe Gly Val Val Arg Ala Cys Arg Glu Asn Ala		
2090	2095	2100
Thr Gly Arg Thr Phe Val Ala Lys Ile Val Pro Tyr Ala Ala Glu		
2105	2110	2115
Gly Lys Arg Arg Val Leu Gln Glu Tyr Glu Val Leu Arg Thr Leu		
2120	2125	2130
His His Glu Arg Ile Met Ser Leu His Glu Ala Tyr Ile Thr Pro		
2135	2140	2145
Arg Tyr Leu Val Leu Ile Ala Glu Ser Cys Gly Asn Arg Glu Leu		
2150	2155	2160
Leu Cys Gly Leu Ser Asp Arg Phe Arg Tyr Ser Glu Asp Asp Val		
2165	2170	2175
Ala Thr Tyr Met Val Gln Leu Leu Gln Gly Leu Asp Tyr Leu His		
2180	2185	2190
Gly His His Val Leu His Leu Asp Ile Lys Pro Asp Asn Leu Leu		
2195	2200	2205
Leu Ala Pro Asp Asn Ala Leu Lys Ile Val Asp Phe Gly Ser Ala		
2210	2215	2220
Gln Pro Tyr Asn Pro Gln Ala Leu Arg Pro Leu Gly His Arg Thr		
2225	2230	2235
Gly Thr Leu Glu Phe Met Ala Pro Glu Met Val Lys Gly Glu Pro		
2240	2245	2250
Ile Gly Ser Ala Thr Asp Ile Trp Gly Ala Gly Val Leu Thr Tyr		
2255	2260	2265
Ile Met Leu Ser Gly Arg Ser Pro Phe Tyr Glu Pro Asp Pro Gln		
2270	2275	2280
Glu Thr Glu Ala Arg Ile Val Gly Gly Arg Phe Asp Ala Phe Gln		
2285	2290	2295
Leu Tyr Pro Asn Thr Ser Gln Ser Ala Thr Leu Phe Leu Arg Lys		
2300	2305	2310
Val Leu Ser Val His Pro Trp Ser Arg Pro Ser Leu Gln Asp Cys		
2315	2320	2325
Leu Ala His Pro Trp Leu Gln Asp Ala Tyr Leu Met Lys Leu Arg		
2330	2335	2340
Arg Gln Thr Leu Thr Phe Thr Thr Asn Arg Leu Lys Glu Phe Leu		
2345	2350	2355
Gly Glu Gln Arg Arg Arg Arg Ala Glu Ala Ala Thr Arg His Lys		
2360	2365	2370
Val Leu Leu Arg Ser Tyr Pro Gly Gly Pro		
2375	2380	

&lt;210&gt; 19

&lt;211&gt; 505

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2190612CD1

&lt;400&gt; 19

Met Glu Gly Gly Pro Ala Val Cys Cys Gln Asp Pro Arg Ala Glu		
1	5	10
Leu Val Glu Arg Val Ala Ala Ile Asp Val Thr His Leu Glu Glu		15
	20	25
		30

Ala Asp Gly Gly Pro Glu Pro Thr Arg Asn Gly Val Asp Pro Pro	35	40	45
Pro Arg Ala Arg Ala Ala Ser Val Ile Pro Gly Ser Thr Ser Arg	50	55	60
Leu Leu Pro Ala Arg Pro Ser Leu Ser Ala Arg Lys Leu Ser Leu	65	70	75
Gln Glu Arg Pro Ala Gly Ser Tyr Leu Glu Ala Gln Ala Gly Pro	80	85	90
Tyr Ala Thr Gly Pro Ala Ser His Ile Ser Pro Arg Ala Trp Arg	95	100	105
Arg Pro Thr Ile Glu Ser His His Val Ala Ile Ser Asp Ala Glu	110	115	120
Asp Cys Val Gln Leu Asn Gln Tyr Lys Leu Gln Ser Glu Ile Gly	125	130	135
Lys Gly Ala Tyr Gly Val Val Arg Leu Ala Tyr Asn Glu Ser Glu	140	145	150
Asp Arg His Tyr Ala Met Lys Val Leu Ser Lys Lys Lys Leu Leu	155	160	165
Lys Gln Tyr Gly Phe Pro Arg Arg Pro Pro Pro Arg Gly Ser Gln	170	175	180
Ala Ala Gln Gly Gly Pro Ala Lys Gln Leu Leu Pro Leu Glu Arg	185	190	195
Val Tyr Gln Glu Ile Ala Ile Leu Lys Lys Leu Asp His Val Asn	200	205	210
Val Val Lys Leu Ile Glu Val Leu Asp Asp Pro Ala Glu Asp Asn	215	220	225
Leu Tyr Leu Val Phe Asp Leu Leu Arg Lys Gly Pro Val Met Glu	230	235	240
Val Pro Cys Asp Lys Pro Phe Ser Glu Glu Gln Ala Arg Leu Tyr	245	250	255
Leu Arg Asp Val Ile Leu Gly Leu Glu Tyr Leu His Cys Gln Lys	260	265	270
Ile Val His Arg Asp Ile Lys Pro Ser Asn Leu Leu Leu Gly Asp	275	280	285
Asp Gly His Val Lys Ile Ala Asp Phe Gly Val Ser Asn Gln Phe	290	295	300
Glu Gly Asn Asp Ala Gln Leu Ser Ser Thr Ala Gly Thr Pro Ala	305	310	315
Phe Met Ala Pro Glu Ala Ile Ser Asp Ser Gly Gln Ser Phe Ser	320	325	330
Gly Lys Ala Leu Asp Val Trp Ala Thr Gly Val Thr Leu Tyr Cys	335	340	345
Phe Val Tyr Gly Lys Cys Pro Phe Ile Asp Asp Phe Ile Leu Ala	350	355	360
Leu His Arg Lys Ile Lys Asn Glu Pro Val Val Phe Pro Glu Glu	365	370	375
Pro Glu Ile Ser Glu Glu Leu Lys Asp Leu Ile Leu Lys Met Leu	380	385	390
Asp Lys Asn Pro Glu Thr Arg Ile Gly Val Pro Asp Ile Lys Leu	395	400	405
His Pro Trp Val Thr Lys Asn Gly Glu Glu Pro Leu Pro Ser Glu	410	415	420
Glu Glu His Cys Ser Val Val Glu Val Thr Glu Glu Glu Val Lys	425	430	435
Asn Ser Val Arg Leu Ile Pro Ser Trp Thr Thr Val Ile Leu Val	440	445	450

Lys	Ser	Met	Leu	Arg	Lys	Arg	Ser	Phe	Gly	Asn	Pro	Phe	Glu	Pro
				455					460					465
Gln	Ala	Arg	Arg	Glu	Glu	Arg	Ser	Met	Ser	Ala	Pro	Gly	Asn	Leu
				470					475					480
Leu	Val	Lys	Glu	Gly	Phe	Gly	Glu	Gly	Gly	Lys	Ser	Pro	Glu	Leu
				485					490					495
Pro	Gly	Val	Gln	Glu	Asp	Glu	Ala	Ala	Ser					
				500					505					

&lt;210&gt; 20

&lt;211&gt; 1572

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477549CD1

&lt;400&gt; 20

Met	Glu	Arg	Arg	Leu	Arg	Ala	Leu	Glu	Gln	Leu	Ala	Arg	Gly	Glu
1				5					10					15
Ala	Gly	Gly	Cys	Pro	Gly	Leu	Asp	Gly	Leu	Leu	Asp	Leu	Leu	Leu
				20					25					30
Ala	Leu	His	His	Glu	Leu	Ser	Ser	Gly	Pro	Leu	Arg	Arg	Glu	Arg
				35					40					45
Ser	Val	Ala	Gln	Phe	Leu	Ser	Trp	Ala	Ser	Pro	Phe	Val	Ser	Lys
				50					55					60
Val	Lys	Glu	Leu	Arg	Leu	Gln	Arg	Asp	Asp	Phe	Glu	Ile	Leu	Lys
				65					70					75
Val	Ile	Gly	Arg	Gly	Ala	Phe	Gly	Glu	Val	Thr	Val	Val	Arg	Gln
				80					85					90
Arg	Asp	Thr	Gly	Gln	Ile	Phe	Ala	Met	Lys	Met	Leu	His	Lys	Trp
				95					100					105
Glu	Met	Leu	Lys	Arg	Ala	Glu	Thr	Ala	Cys	Phe	Arg	Glu	Glu	Arg
				110					115					120
Asp	Val	Leu	Val	Lys	Gly	Asp	Ser	Arg	Trp	Val	Thr	Thr	Leu	His
				125					130					135
Tyr	Ala	Phe	Gln	Asp	Glu	Glu	Tyr	Leu	Tyr	Leu	Val	Met	Asp	Tyr
				140					145					150
Tyr	Ala	Gly	Gly	Asp	Leu	Leu	Thr	Leu	Leu	Ser	Arg	Phe	Glu	Asp
				155					160					165
Arg	Leu	Pro	Pro	Glu	Leu	Ala	Gln	Phe	Tyr	Leu	Ala	Glu	Met	Val
				170					175					180
Leu	Ala	Ile	His	Ser	Leu	His	Gln	Leu	Gly	Tyr	Val	His	Arg	Asp
				185					190					195
Val	Lys	Pro	Asp	Asn	Val	Leu	Leu	Asp	Val	Asn	Gly	His	Ile	Arg
				200					205					210
Leu	Ala	Asp	Phe	Gly	Ser	Cys	Leu	Arg	Leu	Asn	Thr	Asn	Gly	Met
				215					220					225
Val	Asp	Ser	Ser	Val	Ala	Val	Gly	Thr	Pro	Asp	Tyr	Ile	Ser	Pro
				230					235					240
Glu	Ile	Leu	Gln	Ala	Met	Glu	Glu	Gly	Lys	Gly	His	Tyr	Gly	Pro
				245					250					255
Gln	Cys	Asp	Trp	Trp	Ser	Leu	Gly	Val	Cys	Ala	Tyr	Glu	Leu	Leu
				260					265					270
Phe	Gly	Glu	Thr	Pro	Phe	Tyr	Ala	Glu	Ser	Leu	Val	Glu	Thr	Tyr

	275		280		285
Gly Lys Ile Met Asn His Glu Asp His		Leu Gln Phe Pro Pro Asp			
	290		295		300
Val Pro Asp Val Pro Ala Ser Ala Gln		Asp Leu Ile Arg Gln Leu			
	305		310		315
Leu Cys Arg Gln Glu Glu Arg Leu Gly		Arg Gly Gly Leu Asp Asp			
	320		325		330
Phe Arg Asn His Pro Phe Phe Glu Gly		Val Asp Trp Glu Arg Leu			
	335		340		345
Ala Ser Ser Thr Ala Pro Tyr Ile Pro		Glu Leu Arg Gly Pro Met			
	350		355		360
Asp Thr Ser Asn Phe Asp Val Asp Asp		Asp Thr Leu Asn His Pro			
	365		370		375
Gly Thr Leu Pro Pro Pro Ser His Gly		Ala Phe Ser Gly His His			
	380		385		390
Leu Pro Phe Val Gly Phe Thr Tyr Thr		Ser Gly Ser His Ser Pro			
	395		400		405
Glu Ser Ser Ser Glu Ala Trp Ala Ala		Leu Glu Arg Lys Leu Gln			
	410		415		420
Cys Leu Glu Gln Glu Lys Val Glu Leu		Ser Arg Lys His Gln Glu			
	425		430		435
Ala Leu His Ala Pro Thr Asp His Arg		Glu Leu Glu Gln Leu Arg			
	440		445		450
Lys Glu Val Gln Thr Leu Arg Asp Arg		Leu Pro Glu Met Leu Arg			
	455		460		465
Asp Lys Ala Ser Leu Ser Gln Thr Asp		Gly Pro Pro Ala Gly Ser			
	470		475		480
Pro Gly Gln Asp Ser Asp Leu Arg Gln		Glu Leu Asp Arg Leu His			
	485		490		495
Arg Glu Leu Ala Glu Gly Arg Ala Gly		Leu Gln Ala Gln Glu Gln			
	500		505		510
Glu Leu Cys Arg Ala Gln Gly Gln Gln		Glu Glu Leu Leu Gln Arg			
	515		520		525
Leu Gln Glu Ala Gln Glu Arg Glu Ala		Ala Thr Ala Ser Gln Thr			
	530		535		540
Arg Ala Leu Ser Ser Gln Leu Glu Glu		Ala Arg Ala Ala Gln Arg			
	545		550		555
Glu Leu Glu Ala Gln Val Ser Ser Leu		Ser Arg Gln Val Thr Gln			
	560		565		570
Leu Gln Gly Gln Trp Glu Gln Arg Leu		Glu Glu Ser Ser Gln Ala			
	575		580		585
Lys Thr Ile His Thr Ala Ser Glu Thr		Asn Gly Met Gly Pro Pro			
	590		595		600
Glu Gly Gly Pro Gln Glu Ala Gln Leu		Arg Lys Glu Val Ala Ala			
	605		610		615
Leu Arg Glu Gln Leu Glu Gln Ala His		Ser His Arg Pro Ser Gly			
	620		625		630
Lys Glu Glu Ala Leu Cys Gln Leu Gln		Glu Glu Asn Arg Arg Leu			
	635		640		645
Ser Arg Glu Gln Glu Arg Leu Glu Ala		Glu Leu Ala Gln Glu Gln			
	650		655		660
Glu Ser Lys Gln Arg Leu Glu Gly Glu		Arg Arg Glu Thr Glu Ser			
	665		670		675
Asn Trp Glu Ala Gln Leu Ala Asp Ile		Leu Ser Trp Val Asn Asp			
	680		685		690
Glu Lys Val Ser Arg Gly Tyr Leu Gln		Ala Leu Ala Thr Lys Met			

	695	700	705
Ala Glu Glu Leu	Glu Ser Leu Arg Asn Val Gly Thr Gln Thr Leu		
	710	715	720
Pro Ala Arg Pro	Leu Lys Met Glu Ala Ser Ala Arg Leu Glu Leu		
	725	730	735
Gln Ser Ala Leu	Glu Ala Glu Ile Arg Ala Lys Gln Gly Leu Gln		
	740	745	750
Glu Arg Leu Thr	Gln Val Gln Glu Ala Gln Leu Gln Ala Glu Arg		
	755	760	765
Arg Leu Gln Glu	Ala Glu Lys Gln Ser Gln Ala Leu Gln Gln Glu		
	770	775	780
Leu Ala Met Leu	Arg Glu Glu Leu Arg Ala Arg Gly Pro Val Asp		
	785	790	795
Thr Lys Pro Ser	Asn Ser Leu Ile Pro Phe Leu Ser Phe Arg Ser		
	800	805	810
Ser Glu Lys Asp	Ser Ala Lys Asp Pro Gly Ile Ser Gly Glu Ala		
	815	820	825
Thr Arg His Gly	Gly Glu Pro Asp Leu Arg Pro Glu Gly Arg Arg		
	830	835	840
Ser Leu Arg Met	Gly Ala Val Phe Pro Arg Ala Pro Thr Ala Asn		
	845	850	855
Thr Ala Ser Thr	Glu Gly Leu Pro Ala Lys Gly Trp Gly Met Gly		
	860	865	870
Pro Trp Glu Ala	Leu Gly Asn Gly Cys Pro Pro Pro Gln Pro Gly		
	875	880	885
Ser His Thr Leu	Arg Pro Arg Ser Phe Pro Ser Pro Thr Lys Cys		
	890	895	900
Leu Arg Cys Thr	Ser Leu Met Leu Gly Leu Gly Arg Gln Gly Leu		
	905	910	915
Gly Cys Asp Ala	Cys Gly Tyr Phe Cys His Thr Thr Cys Ala Pro		
	920	925	930
Gln Ala Pro Pro	Cys Pro Val Pro Pro Asp Leu Leu Arg Thr Ala		
	935	940	945
Leu Gly Val His	Pro Glu Thr Gly Thr Gly Thr Ala Tyr Glu Gly		
	950	955	960
Phe Leu Ser Val	Pro Arg Pro Ser Gly Val Arg Arg Gly Trp Gln		
	965	970	975
Arg Val Phe Ala	Ala Leu Ser Asp Ser Arg Leu Leu Leu Phe Asp		
	980	985	990
Ala Pro Asp Leu	Arg Leu Ser Pro Pro Ser Gly Ala Leu Leu Gln		
	995	1000	1005
Val Leu Asp Leu	Arg Asp Pro Gln Phe Ser Ala Thr Pro Val Leu		
	1010	1015	1020
Ala Ser Asp Val	Ile His Ala Gln Ser Arg Asp Leu Pro Arg Ile		
	1025	1030	1035
Phe Arg Val Thr	Thr Ser Gln Leu Ala Val Pro Pro Thr Thr Cys		
	1040	1045	1050
Thr Val Leu Leu	Leu Ala Glu Ser Glu Gly Glu Arg Glu Arg Trp		
	1055	1060	1065
Leu Gln Val Leu	Gly Glu Leu Gln Arg Leu Leu Leu Asp Ala Arg		
	1070	1075	1080
Pro Arg Pro Arg	Pro Val Tyr Thr Leu Lys Glu Ala Tyr Asp Asn		
	1085	1090	1095
Gly Leu Pro Leu	Leu Pro His Thr Leu Cys Ala Ala Ile Leu Asp		
	1100	1105	1110
Gln Asp Arg Leu	Ala Leu Gly Thr Glu Glu Gly Leu Phe Val Ile		



1115	1120	1125
His Leu Arg Ser Asn Asp Ile Phe Gln Val Gly Glu Cys Arg Arg		
1130	1135	1140
Val Gln Gln Leu Thr Leu Ser Pro Ser Ala Gly Leu Leu Val Val		
1145	1150	1155
Leu Cys Gly Arg Gly Pro Ser Val Arg Leu Phe Ala Leu Ala Glu		
1160	1165	1170
Leu Glu Asn Ile Glu Val Ala Gly Ala Lys Ile Pro Glu Ser Arg		
1175	1180	1185
Gly Cys Gln Val Leu Ala Ala Gly Ser Ile Leu Gln Ala Arg Thr		
1190	1195	1200
Pro Val Leu Cys Val Ala Val Lys Arg Gln Val Leu Cys Tyr Gln		
1205	1210	1215
Leu Gly Pro Gly Pro Gly Pro Trp Gln Arg Arg Ile Arg Glu Leu		
1220	1225	1230
Gln Ala Pro Ala Thr Val Gln Ser Leu Gly Leu Leu Gly Asp Arg		
1235	1240	1245
Leu Cys Val Gly Ala Ala Gly Gly Phe Ala Leu Tyr Pro Leu Leu		
1250	1255	1260
Asn Glu Ala Ala Pro Leu Ala Leu Gly Ala Gly Leu Val Pro Glu		
1265	1270	1275
Glu Leu Pro Pro Ser Arg Gly Gly Leu Gly Glu Ala Leu Gly Ala		
1280	1285	1290
Val Glu Leu Ser Leu Ser Glu Phe Leu Leu Leu Phe Thr Thr Ala		
1295	1300	1305
Gly Ile Tyr Val Asp Gly Ala Gly Arg Lys Ser Arg Gly His Glu		
1310	1315	1320
Leu Leu Trp Pro Ala Ala Pro Met Gly Trp Gly Tyr Ala Ala Pro		
1325	1330	1335
Tyr Leu Thr Val Phe Ser Glu Asn Ser Ile Asp Val Phe Asp Val		
1340	1345	1350
Arg Arg Ala Glu Trp Val Gln Thr Val Pro Leu Lys Lys Val Arg		
1355	1360	1365
Pro Leu Asn Pro Glu Gly Ser Leu Phe Leu Tyr Gly Thr Glu Lys		
1370	1375	1380
Val Arg Leu Thr Tyr Leu Arg Asn Gln Leu Ala Glu Lys Asp Glu		
1385	1390	1395
Phe Asp Ile Pro Asp Leu Thr Asp Asn Ser Arg Arg Gln Leu Phe		
1400	1405	1410
Arg Thr Lys Ser Lys Arg Arg Phe Phe Phe Arg Val Ser Glu Glu		
1415	1420	1425
Gln Gln Lys Gln Gln Arg Arg Glu Met Leu Lys Asp Pro Phe Val		
1430	1435	1440
Arg Ser Lys Leu Ile Ser Pro Pro Thr Asn Phe Asn His Leu Val		
1445	1450	1455
His Val Gly Pro Ala Asn Gly Arg Pro Gly Ala Arg Asp Lys Ser		
1460	1465	1470
Pro Ser Gln Pro Leu Arg Thr Val Thr Gln Gln Ala Pro Glu Glu		
1475	1480	1485
Lys Gly Arg Val Ala Arg Gly Ser Gly Pro Gln Arg Pro His Ser		
1490	1495	1500
Phe Ser Glu Ala Leu Arg Arg Pro Ala Ser Met Gly Ser Glu Gly		
1505	1510	1515
Leu Gly Gly Asp Ala Asp Pro Thr Gly Ala Val Lys Arg Lys Pro		
1520	1525	1530
Trp Thr Ser Leu Ser Ser Glu Ser Val Ser Cys Pro Gln Gly Ser		

	1535	1540	1545
Leu Ser Pro Ala Thr Ser Leu Met Gln Val Ser Glu Arg Pro Arg			
	1550	1555	1560
Ser Leu Pro Leu Ser Pro Glu Leu Glu Ser Ser Pro			
	1565	1570	

<210> 21  
 <211> 4298  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2564295CB1

<400> 21

```

gccactgaga ggcccaccag gtcccttctt ctcggatctg gcagaccaag gactagcgta 60
cggacctgcg cttaggggtc tccaagagga caaggagcct cctaggagct gagaggggct 120
cccagaggcc aaggctgtcc acgttctccc gggctcaggc tgccagaagt taccgccgag 180
atgtcagagg accgggaagc agaacttaga cacagcgctc ccagaagcc cgggcgcgct 240
ggctgcccct ccggcggtgc agcccactt ggaagaagcc tgtgggctta tcacaccgtt 300
ctccccagag tcaccgggag gagagccggg actggacaca agccagggtt gggacaatgg 360
cagtgcctag tctgtggccc tggggagcat gcctgcctgt gatcttctc tccttgggat 420
ttggcctgga tacagtagag gtgtgcccc gcttgatat tcgctcagag gtggcagagc 480
ttcgtcagct ggagaactgc agcgtggtgg agggccacct gcagatcctg ctcattgtca 540
cagccaccgg ggaggacttc cgcggcctca gcttccctcg cctcaccag gtcaccgact 600
acctgctgct ctcccggtgc tacggactgg agagcctgcg cgacctctc cccaacctag 660
cagtcacccg cgggacgcgc ctcttctctg gctatgact ggtcatctt gagatgccac 720
atctgcgtga cgtggcactg cctgcaactg gggcctgct gcgtggggct gtgcgtgtgg 780
agaagaaaca ggagctctgc cacctctcca ccattgactg .gggactgctg cagccagcac 840
ctggcgccaa ccacatcgtg ggcaacaagc tgggcgagga gtgtgctgac gtgtgccctg 900
gtgtgctggg tgctgctggt gagccctgtg ccaagaccac cttcagcggg cactactgact 960
acagatgctg gacctccagc cactgccaga gagtgtgccc ctgcccccat gggatggctt 1020
gcacagcgag gggcgagtgc tgccacaccg aatgcctggg gggctgcagc cagccagaag 1080
acctcgtgc ctgtgtagct tgccgccacc tctacttcca gggctgctgc ctgtgggctt 1140
gcccgcagag cactaccag tatgagtcc tgcgtgtgt cacagctgag cgctgtgcca 1200
gcctgcactc tgtgcccggc cgtgcctcca ccttcggcat acaccagggc agttgcctgg 1260
cccagtgcct ttctggcttc acccgtaata gcagcagcat attctgccac aagtgcgagg 1320
ggctgtgccc taaagagtgc aaggtaggca ccaagaccat cgactccatc caggcggcac 1380
aggatcttgt gggctgcacg catgtggagg gaagcctcat cctcaacctt cgccagggtt 1440
acaacctgga gccacagctg cagcacagcc tggggctggt agaaaccatt actggcttcc 1500
tcaaaatcaa gcaactcttt gcctcgtgt ccctgggctt tttcaagaac ctcaaaactaa 1560
tccggggaga cgccatggtg gatgggaact acactctcta cgtgctggac aaccagaacc 1620
tacaacagct agggctcctg gtggcccgcg ggtcaccat tcccgtgggc aagatctact 1680
tcgccttcaa cccgcgcctc tgcttgaac acatctaccg actggaggag gtgacaggca 1740
cgcgaggtcg gcagaacaag gctgagatca accccgcac caacggagac cgcgccgctt 1800
gccagactcg caccctgcgc ttctgtgtcca acgtgacgga ggcagaccgc atcctgtctac 1860
gctgggagcg ctatgagcca ctggaggccc gcgacctgct cagcttcac gtgtactaca 1920
aggagtcccc attccagaac gccacagagc acgtgggtcc agatgcttgt ggaaccaga 1980
gctggaacct gctggatgtg gagctgcccc taagccgcac ccaggagcca ggggtgacct 2040
tagcctccct caagccttgg acacagtacg cagtgtttgt gcgggccatc acgctaacca 2100
ctgaggagga cagccctcat caaggagccc agagtcccat cgtctacctc cgaacgctgc 2160
ctgcagctcc caccgtgccc caagacgtca tctccacgtc caactcctcc tcccacctcc 2220
tggtgcgtg gaagccaccg acccagcgca atgggaacct cacctactac ctggtgctgt 2280
ggcagcggtt ggcagaggac ggcgacctct acctcaatga ctactgccac cgcggcttgc 2340
ggctgcccac cagcaacaac gatccgcgct tcgacggcga agacggggat cctgaggccg 2400

```

```

agatggagtc cgactgctgc ccttgccagc acccacctcc tggtcaggtt ctgccccgc 2460
tggaggcgca agaggcctcg ttccagaaga agtttgaaaa ctttctacac aacgcgatca 2520
ccatccccat atccccctgg aaggtgacgt ccatcaacaa gagccccaa agggactcag 2580
ggcggcaccc cgggcgagct gggccccctc ggctgggggg caacagctcg gatttcgaga 2640
tccaggagga caaggtgccc cgtgagcgag cggtgctgag cggcctgcgc cacttcacgg 2700
aataccggat cgacatccat gcctgcaacc acgcgcgca caccgtgggc tgcagcgccg 2760
ccaccttcgt ctttgcgcg accatgcccc acagagaggc tgatgggtatt ccaggaaagg 2820
tggcctggga ggcctccagc aagaacagtg tccttctgcg ctggctcgag ccaccagacc 2880
ccaacggact catcctcaag tacgaaatca agtaccgccc cttgggagag gaggccacag 2940
tgctgtgtgt gtcccgtctt cgatatgcca agtttggggg agtccacctg gccctgctgc 3000
cccctggaaa ctactctgcc agggttaggg caacctcact ggctggcaat ggctcttgga 3060
cagacagtgt tgcctctac atccttggcc cagaggagga ggatgctggg gggctgcatg 3120
tctctctcac tgccaccct gtggggctca cgctgctcat cgttcttgct gcccttggtt 3180
tcttctacgg caagaagaga aacagaacct tgtatgcttc tgtgaatcca gagtacttca 3240
ggcctctga tatgtatgct cctgatgaat gggaggtgcc tcgggagcag atctcgataa 3300
tccgggaact gggccagggc tcttttgga tggatatga ggggctggca cgaggacttg 3360
aggctggaga ggagtccaca cccgtggccc tgaagacggg gaatgagctg gccagcccac 3420
gggaatgcat tgagttcctc aaggaagctt ctgtcatgaa agccttcaag tgtcaccatg 3480
tgggtgcgtct cctgggtgtg gtatctcagg gccagccaac tctggtcatc atggagttaa 3540
tgaccctggg ggacctcaag agccatcttc gatctttgag gcctgaggca gagaacaacc 3600
ctgggctccc acagccagca ttgggggaaa tgatccaaat ggctgggtgag attgcagacg 3660
gcatggccta ccttgcctgc aacaagtttg tgcaccgaga tctagcagcc cgcaactgca 3720
tgggtgtccca ggacttcacc gtcaagatcg gggacttcgg gatgactcgg gacgtgtatg 3780
agacagacta ttaccgcaag ggtgggaagg ggctgctgcc cgtgcgctgg atggccccg 3840
agtccctcaa agatgggagc ttcaccaccc actcggatgt ctggctcctt ggcgtggtac 3900
tctgggagat tgtgacctg gcagaacaac cctaccaggg cctgtccaat gaggaggtgc 3960
tgaagtctgt catggatggc ggggtcctgg aggagctgga gggctgtccc cttcagctgc 4020
aggagctgat gagccgctgc tggcagccga acccagcct gcgcccctt ttcacacaca 4080
ttctggacag catacaggag gagctgcggc cctccttcg cctcctctcc ttctactaca 4140
gcccggatg cggggggccc cggggctccc tgctaccac cgatgcagag cctgactcct 4200
caccactcc aagagactgc agccctcaaa atgggggtcc agggcactga ggggcacctc 4260
attccctggc tggcctccca tggggagaca ggaaggga 4298

```

&lt;210&gt; 22

&lt;211&gt; 2863

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2837050CB1

&lt;400&gt; 22

```

atgatggaag aattgcatag cctggacca cgacggcagg aattattgga ggccaggttt 60
actagagtag gtgttagtaa gggaccactt aatagtgagt cttccaacca gagcttgtgc 120
agcgtcggat ccttgagtga taaagaagta gagactcccg agaaaaagca gaatgaccag 180
cgaaatcgga aaagaaaagc tgaaccatat gaaactagcc aagggaagg cactcctagg 240
ggacataaaa ttagtgatta ctttgagcga cgagtagaac agcccctcta tggtttagat 300
ggcagtgtctg caaaggaggc aacggaggag cagtctgctc tgccaaccct catgtcagtg 360
atgctagcaa aacctcggct tgacacagag cactgtggcg aaaggggagc tggcctctgc 420
ttcacttttg tttcagctca gcaaaacagt ccctcatcta cgggatctgg caacacagag 480
cattcctgca gctcccaaaa acagatctcc atccagcaca gacagacca gtccgacctc 540
acaatagaaa aaatatctgc actagaaaac agtaagaatt ctgacttaga gaagaaggag 600
ggaagaatag atgatttatt aagagccaac tgtgatttga gacggcagat tgatgaacag 660
caaaagatgc tagagaaata caaggaacga ttaaatagat gtgtgacaat gagcaagaaa 720
ctccttatag aaaagtcaaa acaagagaag atggcgtgta gagataagag catgcaagac 780

```

```

cgcttgagac tggggcactt tactactgtc cgacacggag cctcatttac tgaacagtgg 840
acagatgggt atgcttttca gaactttatc aagcaacagg aaaggataaa ttcacagagg 900
gaagagatag aaagacaacg gaaaatgtta gcaaagcgga aacctcctgc catgggtcag 960
gccccctcctg caaccaatga gcagaaacag cggaaaaagca agaccaatgg agctgaaaat 1020
gaaacgttaa cgtttagcaga ataccatgaa caagaagaaa tcttcaaact cagattaggt 1080
catcttaaaa aggaggaagc agagatccag gcagagctgg agagactaga aagggttaga 1140
aatctacata tcagggaact aaaaaggata cataatgaag ataattcaca atttaaagat 1200
catccaacgc taaatgacag atatttgttg ttacatcttt tgggtagagg aggtttcagt 1260
gaagtttaca aggcatttga tctaacagag caaagatacg tagctgtgaa aattcaccag 1320
ttaataaaaa actggagaga tgagaaaaag gagaattacc acaagcatgc atgtagggaa 1380
taccggattc ataaagagct ggatcatccc agaatagtta agctgtatga ttacttttca 1440
ctggatactg actcgttttg tacagtatta gaatactgtg agggaaatga tctggacttc 1500
tacctgaaac agcacaactt aatgtcagag aaagaggcct ggtccattat catgcagatt 1560
gtgaatgctt taaagtactt aaatgaaata aaacctcca tcatacacta tgacctcaa 1620
ccaggtaata ttcttttagt aaatggtaca gtgtgtggag agagaaaaat tacagatttt 1680
ggtctttcga agatcatgga tgatgatagc tacaattcag tgggtggcat ggagctgaca 1740
tcacaagggtg ctggcactta ttggtattta ccaccggagt gttttgtggt tgagaaagaa 1800
ccaccaaaaga tctcaaataa agttgatgtg tggtcggtgg gtgtgatctt ctatcagtgt 1860
ctttctggag ggaagccttt tggccataac cagtctcagc aagacatcct acaagagaat 1920
actattctta aagctgctga agtgagttc ccgcaaagc cagtagtaac acctgaagca 1980
aaggcgttta ttcgacgatg cttggcctac cgaaaggagg actgcattga tgcccagcag 2040
ctggcctgtg atccctactt gttgcctcac atccgaaagt cagtctctac aagtagccct 2100
gctggagctg ctattgcac aacctctggg gcgtccaata acagttcttc taattgagac 2160
tgactccaag gccacaaact gttcaacaca cacaaagtgg acaaatggcg ttcagcagcg 2220
ggtttggaac atagcgaatc tgaatggatc tgatgaaacc tgaaccaggt gcttttattt 2280
tettgctttt ttcccatcca ctgagcatga cagcatggat tctctttaag gagaaacctt 2340
gggcagctcc agccaggcct cataggaaaa ggcccgcat gaggttctgg cgtcaatggc 2400
cactgtgtat ggctgctctg agtgaggaaa aaactaaaaa gaaaaactgg ttccatgtac 2460
tgtgaacttg aaaacatgca gactcacggg ggttctctgat gcaatgcttc agatgaagat 2520
tgtggacttg aaaatacaga ctagaaggcc gggcacagtg gctcatgcct gtaatctcag 2580
cactttggga ggccaaggaa ggtggatcac aaggtcagga gatcgagacc atcctggcta 2640
acacagtga acccgtctc tactaaaaat acaaaaaaat tagccaggct tgggtggggg 2700
cgcctatagt ccagctact tgggagactg aggcaggaga atgtcgtgaa cccgagaggc 2760
ggagcttgca gtgagcag atcacgccac tgcactccag cctgggagac agagtgcac 2820
tccgtctcaa aaaataata tataataaaa taataaaaaa aaa 2863

```

<210> 23

<211> 1494

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474590CB1

<400> 23

```

atgtactctg acagcgagga tgagtcacaa gagctcagca ctgtgctcag catgtttgag 60
gagaaggagt tcaccaggca gtacaccgtc ctgaagacct tgagccagca tggcactact 120
gaagtgaggc tatgtctcca tcacctcaca ggtgtcacag ttgtgtcaa agctctgaag 180
taccagaggt ggtgggagcc aaaggtttca gaagtcgaaa tcatgaagat gtcagccac 240
cctaacattg ttcccttct gcaagtgata gagacagaac agaacattta tctgattatg 300
gaagtggccc aaggcacaca gctacataat cgagtccagg aggttaggtg cctgaaggaa 360
gatgaagcaa gaagcatatt tgttcagttg ctcagtgcc taggctactg tcatggtgaa 420
ggtgtgttgc acagagacct aaagcctgac aatgtcatag ttgatgagca tggaaatgtc 480
aaaattgttg actttgggct aggtgccaga ttcatgcctg ggcagaaatt ggaaaggctg 540
tgtggagcct tccagttcat tctccagag atattcctag ggctccctta tgatggccca 600

```

```

aaagtagaca tatgggcctt ggggggttctt ttgtattata tgggtgacagg gattttttcca 660
tttgtagggt ccaccttggtc agaaattagc aaggaagttc tacaaggag gtatgaaatt 720
ccttataatc tctctaaaga ctttaaggagc atgataggcc tgttattggc aacaaacgca 780
aggcagaggc caactgcaca agacctccta agtcatccat ggcttcagga aggggaaaag 840
actatcacat ttcattccaa tggagacacc agctttccag accctgacat aatggcagcc 900
atgaaaaata ttgggtttca tgtgcaggac attagagaat cattaaaaca cagaaagttc 960
gatgaaacta tggctacata taacttactg agagctgagg catgtcagga tgatggcaat 1020
tatgttcaaa caaagttaat gaatccaggg atgccacctt tcccttcagt aacagactct 1080
ggagcttttt ctctgcctcc taggagaagg gccagtgaac ctctctttaa agtattagtc 1140
tcatctactg aagaacatca attaagacaa actgggggga caaatgcccc ttttccacc 1200
aagaaaacac ccactatggg cagaagtcag aaacagaaac gtgccatgac tgccccctgt 1260
atgtgtttac tgagaaacac ttacatagat acagaagaca gcagcttttg cactagctcc 1320
caggcagaaa agacttcaag tgatccagag aaaagtgaga cttcaacttc atgccctctg 1380
acacctaggg gctggaggaa atggaagaag agaattgtag catgcatcca gacattgtgt 1440
tgctgcacgt tgcctcaaaa aaaatgtccg aggagtgtgc atccccaaaa gtga 1494

```

<210> 24

<211> 2341

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474594CB1

<400> 24

```

atgtcagggc tgggtgctgat gctggcgggc cggtgcattg tgggcagctc cccgctctgc 60
cgctgccgcc gctgctgccc aaggaggatc ggggcccggc cgggccggga tgatccgggt 120
cggaaggccg ccgcccgggg agggagcggg tcacccaacg ccgactgag cgcggccgc 180
cccgcccggg ccccggggga tgcgcgccc cgagctgctg cctccgccc cgcgcagcc 240
gcagccgagc cgggcacaga gcaggtagat ggccccctca gggcaggccc ggcgacacc 300
cctccctctg gctggcggat gcagtgccta gcggccgccc ttaaggacga aaccaacatg 360
agtgggggag gggagcaggc cgacatcctg cgggccaaact acgtggtcaa ggatcgctgg 420
aaggtgctga aaaagatcgg gggcgggggc tttggtgaga tctacgaggc catggacctg 480
ctgaccaggg agaattgtgg cctcaagggt gagtcagccc agcagcccaa gcaggtctc 540
aagatggagg tggccgtgct caagaagtgt caagggaagg accatgtgtg caggttcatt 600
ggctgtggca ggaacgagaa gtttaactat gtagtgatgc agctccagg cgggaacctg 660
gccgacctgc gccgtagcca gccgcgaggc accttcacgc tgagcaccac attgcggtg 720
ggcaagcaga tcttgagtc catcgaggcc atccactctg tgggcttct gcaccgtgac 780
atcaagcctt caaactttgc catgggcagg ctgcccctca cctacaggaa gtgctatatg 840
ctggacttcg ggctggcccc gcagtacacc aacaccacgg gggatgtgag gccccctcgg 900
aatgtggccg ggtttcgagg aacggttcgc tatgcctcag tcaatgccc caagaaccgg 960
gagatgggcc gccacgacga cctgtggtcc ctcttctaca tgcgtggtgga gtttgagtg 1020
ggccagctgc cctggaggaa gatcaaggac aaggaacagg tagggatgat caaggagaag 1080
tatgagcacc ggatgctgct gaagcacatg ccgtcagagt tccacctct cctggaccac 1140
attgccagcc tgactactt caccaagccc gactaccagt tgatcatgtc agtgtttgag 1200
aacagcatga aggagaggga cattgccgag aatgagccct ttgactggga gaaggcagg 1260
accgatgccc tctgtccac gagcacctct accccgcccc agcagaacac ccggcacagc 1320
gcagccatgt ttggggtggt caatgtgacg ccagtgcctg gggacctgct ccgggagaa 1380
accgaggatg tgctacaggg agagcacctg agtgaccagg agaatgcacc cccaattctg 1440
ccggggaggc cctctgaggg gctgggcccc agtccccacc ttgtcccca ccccggggt 1500
cctgaggctg aagtctggga ggagacagat gtcaaccgga acaaactcc gatcaacatc 1560
ggcaaagtaa ctgccgccag ggcaaggggc gtgggtggcc tttctctca ccccgattc 1620
ccagccttgt gcccctgccc tgttcctct aagcacctg tccccggcca tctgctgct 1680
tgccctgct ctgtttcccg gtcctcccc gactagcct cgtgtgtct tccatcatca 1740
tcatctctg tctccttcac cctgaggaga ccatccgcc acagccgct catcagcccc 1800

```

```

agctcatggc actccccctct cctgcagagc ccctgtgtgg aggaggaaca gagccgaggc 1860
atgggggtgcc ccagctcccc agtgcggtgcc ccccagact ccccacaaac ccagtcctgt 1920
tctctgcgct accggagggt gaacagccct ggtcagaaa ggctgtccac ggcggacggg 1980
cgagtgggagc tacctgagag gaggtgggtc tggggccagg ggcattggtg gggcccaagg 2040
ccctctccgc cttcacgtgg ctggtctgga ggaaaagtta gatgtgtggc ggaggtgggc 2100
agaccctggg aagtgtctgag agggttatac ttgggcctgg ggtcagactc agttggggcc 2160
agagacaggg cctgggagaa ccagtggggg atccagagag gtcccggtc atgccaggaa 2220
acgtaattgg gtgagtgcag gctgcaggag ggacaggtgg ggcgcctggg cccaggaagg 2280
gtgaggggag agttggttgg tgggtgtgtg tgcttcaga atctctctc ctagagacta 2340
a
2341

```

&lt;210&gt; 25

&lt;211&gt; 2552

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477585CB1

&lt;400&gt; 25

```

cgcggtgtct ggcgtcgggt ggggtgtggtt gcccctagtt tgaggcctgc ccgattaccc 60
gcaagacttg ggcagccccg ggcgcgcgtc cgaccacgac agggaaagga acctaatct 120
catctttaaataaaggagaa ttactgagtg acctgaagga cccttttcag ctggaaagtc 180
tgaactgacc aacactggat gaatttgacc atttcttagg agactggaat gttaagtctc 240
tataaatgaa tgaaccagtt ctctcttgtt tggagcaatg ctgaaattcc aagaggcagc 300
taagtgtgtg agtggatcaa cagccatttc cacttatcca aagacctga ttgcaagaag 360
atacgtgctt caacaaaaac ttggcagtggt aagttttgga actgtctatc tggtttcaga 420
caagaaagcc aaacgaggag aggaattaaa ggtacttaag gaaatatctg ttggagaact 480
aaatccaat gaaactgtac aggccaatat ggaagcccaa ctctctcca agctggacca 540
cccagccatt gtcaagttcc atgcaagttt tgtggagcaa gataatttct gcattatcac 600
ggagtactgt gagggccgag atctggacga taaaattcag gaatataaac aagctggaaa 660
aatctttcca gaaaatcaaa taatagaatg gtttatccag ctgtgtctgg gagttgacta 720
catgcatgag aggaggatac ttcatcgaga cttaagtc aagaatgtat ttctgaaaaa 780
taatctcctt aaaattggag attttggagt ttctcgactt ctaatggggg cctgtgacct 840
ggccacaact ttaactggaa ctccccatta tatgagtcct gaggtctctga aacaccaagg 900
ctatgacaca aagtcggaca tctggtcact ggcattgcat ttgtatgaga tgtgtgcat 960
gaatcatgca ttgcgtggct ccaatttctt atccattgtt ttaaaaaattg ttgaagggtga 1020
cacaccttct ctccctgaga gatatccaaa agaactaat gccatcatgg aaagcatgtt 1080
gaacaagaat ccttcattaa gaccatctgc tatcgaaatt ttaaaaaatcc cttacctga 1140
tgagcagcta cagaacctaa tgtgtagata ttcagaaatg actctggaag acaaaaaattt 1200
ggattgtcag aaggaggctg ctcatataat taatgccatg caaaaaagga tccacctgca 1260
gactctgagg gcaactgtcag aagtacagaa aatgacgcca agagaaagga tgcggctgag 1320
gaagctccag ggcgtgatg agaaagccag gaagctgaaa aagattgtgg aagaaaaata 1380
tgaagaaaat agcaaacgaa tgcaagaatt gagatctcgg aactttcagc agctgagtgt 1440
tgatgtactc catgaaaaaa cacattttaa aggaatggaa gaaaaggagg agcaacctga 1500
gggaagactt tcttgttcac ccaggacga ggatgaagag aggtggcaag gcagggaaga 1560
ggaatctgat gaaccaactt tagagaacct gcctgagtct cagcctattc cttccatgga 1620
cctccacgaa cttgaatcaa ttgtagagga tgccacatct gacctggat accatgagat 1680
cccagaagac ccacttgtgg ctgaagagta ctacgtgat gcatttgatt cctattgtgt 1740
agagagtgat gaggaggaag aagaaatagc gttagaaaga ccagagaaag aaatcaggaa 1800
tgagggatcc cagcctgctt acagaacaaa ccaacaggac agtgatatcg aagcgttggc 1860
caggtgtttg gaaaatgtcc tgggttgac ttctctagac acaaagacca tcaccacat 1920
ggctgaagac atgtcccgag gaccaccaat tttcaacagt gtgatggcca ggaccaagat 1980
gaaacgcatg agggaatcag ccatgcagaa gctggggaca gaagtatttg aagaggtcta 2040
taattacctc aagagagcaa ggcacagaa tgctagcgaa gcagagatcc gcgagtgtt 2100

```

```

ggaaaaagtgtgtgcctcaagccagcgactgttttgaagtggaccagctccgttacttttga 2160
agagcagttgtctgatcacgagtgggaaaagaacctactctccagaaccatctctaggccaac 2220
tatcaaaaagaaagcagaagtccaagtggacaaatttatgtgaaaattcatttaacatata 2280
agctgaactctattatgggggaatggatacaaaagcagagctcccatcttgactttcaatt 2340
cctcatcagagagtactgggtctcttttagagagtagtaagcattggctgcctatgcttggagt 2400
cataagtgttatttggactataccctgagataagcttatagatcaagtttggctcccttg 2460
aaaagcatttctctcatgtgcgacctcagggtctccagcaggattgagtcacctgacga 2520
tgaccggggaagaagccgtgtgctcttcattat2552

```

&lt;210&gt; 26

&lt;211&gt; 2176

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477587CB1

&lt;400&gt; 26

```

ctcaccgccc tccccaccag cccagcagc acaaacgctg ggcttctctg ccgctccgaa 60
ttctgctgtg gctccccagt gccaggggcc atcaagcccc aggttttcat gtgggcccga 120
gggaccgccc agcagggagg acccgaaatg gcacacattc agaacgtcga ggctcatacc 180
tcaagcgcat tgtgggggag aagccccggg aagcccccaa cccccacgc gcgagagagc 240
ctcagtttcc cgctcgagcg gccccggagc ggccgcagcg cgttggtctc ggcccggtcg 300
cgccagagtc cgcgcatgga gccccggccg cggcgggcggc gcaggagtcg cccctgggtc 360
gccgccttcc tgcgagaccc gggctcgggc cgcgtgtaca ggcgcgggaa gctgatcggc 420
aagggcgctt tcagccgctg ctacaagctg acagacatgt ccaccagcgc cgtgttcgcc 480
ctcaaggtgg tgccgtgtgg cggggctggg gccgggtggc ttcgcccga gggaaaggtg 540
gagcgtgaga ttgccctgca tagccgcctg cgaccccgca acatcgtggc ttccacgga 600
cactttggtg accgcgacca cgtgtacatg gtgctggagt actgcagccg ccagtctttg 660
gcccagctgc tgagggcgcg gcagatcctg acggagccag aagtgcgcga ctacctgcg 720
ggcctggtca gcggtctgcg ctacctgcac cagcgggtgca tcctgcaccg cgacctgaag 780
ctcagtaact tcttctttaa caagaacatg gaggtgaaga ttggagacct gggactggcg 840
gccaaggttg ggcaggggg cgctgccac aggtacacgg tgcgtactgg caccaccacc 900
ttcatggcct caccctgtc ggagatgtac caaaacatcc gtgagggcca ctaccgccaa 960
cccgtcacc tgtctgcaa tgcgcggcgc ctcatcgtgc acctcctagc acccaaccgg 1020
gccgagcggc ccagcctgga ccacctgctg caggacgact tcttcacaca gggtttact 1080
ccagaccggc tgccggccca ctctgccac agtcccccca tcttgccat accccgcct 1140
ctgggcagga tcttcggaa ggtgggccag cggctgctca ccagtgccg gccaccctgc 1200
cccttcacgc ctaaagaggc ctcggtcca ggagaaggtg ggccagacc tgactccatg 1260
gagtgggacg gcgagagtc cctgtctgcg aaagaggttc cctgcctgga aggccccatc 1320
cacctggtcg cacaaggag cctgcagagt gacctggccg ccacacagga cccctggga 1380
gagcagcagc ccacctctg gggcccaaa tgggtggatt attccagcaa ataggcttt 1440
ggctaccagc tcttggacgg gggcgcacg ggacggcacc cacatggccc tgcgacccc 1500
cggaggtatt tattaagcac ctactgtgca cacctacagg tgcctcctgc ctgccaagt 1560
tgctacatgc ccaactgcgg gaggtggaa gccttcgccc tgagggatgt gcccggtcg 1620
ctgggcgcca agctggccgt gctgcagctc tttgcgggt gcctgcggcg gcggctgcgg 1680
gaggagggga cctccccac acctgtgcca cctgctggac ccggcctctg cctcctgcgc 1740
ttcctggcct ctgagcagc cctgctgctg ctgttcagca atgggatggt gcaggtgagc 1800
ttcagtgagg tcccgccca actggtgctg agtggcagg gtgagggttt gcagctcacc 1860
ctctgggagc aggggtcccc tggcacctcc tactccctgg acgtcccgcg gagccacggc 1920
tgcccccca ccaccggaca gcaccttcac cagccctcc gcctgctgca gagtatctag 1980
tgccctgag ggtcagagt gaccctgca tggtagtgcc agggaccag gctccatttc 2040
cattcctgtg gctccccag aggggctgtc ctgggggaga gctggggggc acacgggagg 2100
tgggttcttg ccttgtggca tgactgttca acccagactt tgctgggatc tcttctttt 2160
tcattaaaga caattc2176

```

&lt;210&gt; 27

&lt;211&gt; 4277

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7594537CB1

&lt;400&gt; 27

```
tgtgagcaga gttcttgaag ctccactcct cctggggaag ccgagctgtg tgggagcctt 60
cttactgtgc cgggagcgtg tgaattggaa aggatcctga gaactggcta gtcccagttc 120
ctctccggaa aagcagtggc tctcgcttca gagatgcgtc cagctttcgc ctgcatcaca 180
ctgcattctt taacacatta ttgaaattac aagcatccaa agcagtttca tgtggacaga 240
ttgcatattt tgaagcctg aggtatttta tcatgaaaca tgccatgtgg aatctttgaa 300
gcatagacct ctgcgcaaca cctgaataaa gaatctttta cctggtatgt gacagagctt 360
ctcaccacca ccatgacaaa ccaggaaaaa tgggcccacc tcagcccttc ggaatttttc 420
caacttcaga aatatgctga gtattctaca aagaaattaa aggatgttct tgaagaattc 480
catggtaatg gtgtgcttgc aaagtataat cctgaaggga caatagattt tgaaggtttc 540
aaactattca tgaagacatt cctggaagcc gagcttcctg atgatttcac tgcacacctt 600
ttcatgtcat ttagcaacaa gtttcctcat tctagtccaa tggtaaaaag taagcctgct 660
ctcctatcag gcggtctgag aatgaataaa ggtgccatca cccctccccg aactacttct 720
cttgcaaaaa cgtgttcccc agaagtaate catctgaagg acattgtctg ttacctgtct 780
ctgcttgaaa gaggaagacc tgaggataag cttgagtta tgtttcgctt ttatgacacg 840
gatgggaatg gcttcctgga cagctcggag ctagaaaaa tcatcagtc gatgatgcat 900
gttgacagaat accttgagtg ggatgtcact gaacttaatc caatcctcca tgaatgatg 960
gaagaaattg actatgatca tgatggaacc gtgtctctgg aggaatggat tcaaggagga 1020
atgacaacga ttccacttct tgtgtcctg ggcttagaaa ataacgtgaa ggatgatgga 1080
cagcacgtgt ggcgactgaa gcactttaac aaacctgcct attgcaacct ttgcctgaac 1140
atgctgattg gcgtggggaa gcagggcctc tgctgttcct tctgcaagta cacagtccat 1200
gagcgtgtg tggctcgagc acctccctct tgcacaaaga cctatgtgaa gtccaaaagg 1260
aacactgatg tcatgcacca ttactgggtt gaaggtaact gcccaaccaa gtgtgataag 1320
tgccacaaaa ctgttaaatg ttaccagggc ctgacaggac tgcatttgtt ttggtgtcag 1380
atcacactgc ataataaatg tgcttctcat ctaaaacctg aatgtgactg tggacctttg 1440
aaggaccata ttttaccacc cacaacaatc tgtccagtgg tactgcagac tctgccact 1500
tcaggagttt cagttcctga ggaaagacaa tcaacagtga aaaaggaaaa gagtgttcc 1560
cagcagccaa acaaagtgat tgacaagaat aaaatgcaaa gagccaactc tgttactgta 1620
gatggacaag gcctgcaggt cactcctgtg cctgggtactc acccactttt agttttgtg 1680
aaccacaaaa gtggtggaaa acaaggagaa cgaatttaca gaaaattcca gtatctatta 1740
aatcctcgtc aggtttacag tctttctgga aatggaccaa tgccagggtt aaactttttc 1800
cgtgatgttc ctgacttcag agtggttagc tgtggtggag atggaaccgt gggctgggtt 1860
ttggattgca tagaaaaggc caatgtaggc aagcatcctc cagttgcgat tctgcctctt 1920
gggactggca atgatctagc aagatgcctg cgatggggag gaggttacga aggtgagaat 1980
ctgatgaaaa ttctaaaaga cattgaaaac agcacagaaa tcatgttgga caggtggaag 2040
tttgaagtca tacctaataa caaagatgag aaaggagacc cagtgcctta cagtatcatc 2100
aataattact tttccattgg cgtggatgcc tccattgcac acagattcca catcatgaga 2160
gaaaaacacc cagagaaatt caacagtaga atgaagaaca aatttttgta ttttgagttt 2220
ggcacatctg aaactttctc agccacctgc aagaagctac atgaatctgt agaaatagaa 2280
tgtgatggag tacagataga tttaataaac atctctctgg aaggaattgc tattttgaat 2340
ataccaagca tgcatggagg atccaatctt tggggagagt ctaagaaaag acgaagccat 2400
cgacgaatag agaaaaaagg gtctgacaaa aggaccaccg tcacagatgc caaagagttg 2460
aagtttgcaa gtcaagatct cagtgaccag ctgctggagg tggctcggtt ggaaggagcc 2520
atggagatgg ggcaaatata cacaggcctg aaaagtgtg gccggcggct ggctcagtgc 2580
tcctgcgtgg tcatcaggac gagcaagtct ctgccaatgc aaattgatgg ggagccattg 2640
atgcagacc catgcacaat aaaaattaca cacaagaacc aagcccaat gctgatggg 2700
ccgcctccaa aaaccgggtt attctgtctc ctcgtcaaaa ggacaagaaa ccgaagcaag 2760
```



```

gaataatcct gtgttggttc actcttagaa attgaattag cataattggg ccatggaaca 2820
catatgctgg aaatctttga accatttcaa gtctcctgct catgcaaaat catggaagtg 2880
gtttaacagt ttttggtact aagctaattgt aaaattcagc tattagaaaa tttattgtct 2940
cagtttttat aggcatcttt gcatgaagaa agcagaagtt tacctgaagt gatactgcat 3000
atTTTTggtg catgcattcc catagatttt tacatctccc acccaactct tccccatttt 3060
ccttttacta acctgtgaga aaaacccgtg aaacatgaaa aaggaaatac catgggaaac 3120
gtgattctca gtgtgattcc aattattacg aagcactaat cagtaacgct acaatgatca 3180
taattgcaga ttgctatacg tttccctttt agaatcagtg tatcagtgac ctatgacttg 3240
aggagaaact ttttaattcg agattttatt aaatagttga ctacaatacc ttgctatata 3300
tacatagttt ttcttcaaca tcttaactct tctgagtggg aataaaaaata tcaggcataa 3360
ggttttctca tgctgaaaaa tagaacgcgg tttttatttt gcttagtttt ctttttaatt 3420
ccagaaataa gtgaaaacat gttacttgac agtcaagtgt ggtaatatgg caagccttgt 3480
tcctttctgc atgagaatct aggagagaat tcataaccac accaataacg aaatagaagt 3540
tttaaaactat gtgcctaata aatgtgtttc ccaccaaaga ttcagaaaaa aatgcttgag 3600
agaaatgggt taatgcataa ttaattaagc attgtggagc aaatttaggg ttctgtgat 3660
taattttgtg atgactaaaa tgctggaaag caagtgagtt gccattaat tatgattaaa 3720
attctcacct ttcacagaca gacaataagc cagacaacac aatcaaagct caatagatga 3780
tttcttgctt ttttcagtca tttataaata taggtgtaat tttcatgga tcagttaagt 3840
acacttgaag gaagtaaagt attgtatcag tttatttcta gtataaatgg gtacctgtaa 3900
taatactgag ctcttggaag cgaatcatgc atgcaattag ctccctctc ctcacctact 3960
ccactcccat ctttatgaca tttcaaagt ttatttggaa acaacagcct agatcactgt 4020
tgaagggtgt catggcatag ttggagtctc tgactgttta aagaaatcac agaacagtc 4080
ttttctttta gtgtttcatt aagcctatga tgtaaaatga aatgcttctg agcagcttg 4140
taatatgttt cattcatatt gacctgcac tcatcattgc atgttttatg ttttcaaaca 4200
tgccataagg aaaacgagtg cctgaactgc atgatttatt agtttctctc cactctgcat 4260
taaagtgcata atgattt
4277

```

&lt;210&gt; 28

&lt;211&gt; 2616

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70467491CB1

&lt;400&gt; 28

```

atgtccacta ggacccatt gccaacggtg aatgaacgag acactgaaa cgctgtattg 60
ccgcacacgt cacatggaga tgggcgtcaa gaagttacct ctcgtagcag ccgctcagga 120
gctcgggtga gaaactctat agcctcctgt gcagatgaac aacctcacat cggaaactac 180
agactgttga aaacaatcgg caagggaat tttgcaaaag taaaattggc aagacatc 240
cttacaggca gagagaaaaa tgtagaata tccaaagaaa ttgataattt tctaggaaaa 300
catgacttac caaaattaac tctagaaaag aatcgatata catcagtaac aacagaagt 360
gagaaagtag ttaacatatt gccaaacctg gaattcatga ttgaattctt tgagatctac 420
tctatagggtg aagtatttga ctatttgggt gcacatggca ggatgaagga aaaagaagca 480
agatctaaat ttagacagat tgtgtctgca gttcaatact gccatcagaa acggatcgta 540
catcgagacc tcaaggctga aaatctattg ttagatgccg atatgaacat taaaatagca 600
gatttcggtt ttagcaatga atttactgtt ggcggtaaac tcgacacggt ttgtggcagt 660
cctccatacg cagcacctga gctctccag ggcaagaaat atgacgggcc agaagtggat 720
gtgtggagtc tgggggtcat tttatacaca ctagttagtg gctcacttcc ctttgatggg 780
caaaacctaa aggaactgag agagagagta ttaagaggga aatacagaat tcccttctac 840
atgtctacag actgtgaaaa ccttctcaaa cgtttcctgg tgctaaatcc aattaaacgc 900
ggcactctag agcaaatcat gaaggacagc tggatcaatg cagggcatga agaagatga 960
ctcaaaccat ttgttgaaac agagctagac atctcagacc aaaaaagaat agatattatg 1020
gtgggaatgg gatattcaca agaagaaatt caagaatctc ttagtaagat gaaatagcat 1080
gaaatcacag ctacatattt gttattgggg agaaaatctt cagagctgga tgctagtgat 1140

```

```

tccagttcta gcagcaatct ttcacttgct aaggtttaggc cgagcagtgga tctcaacaac 1200
agtactggcc agtctcctca ccacaaagtg cagagaagtg tttcttcaag ccaaaagcaa 1260
agacgctaca gtgaccatgc tggaccagct attccttctg ttgtggcgta tccgaaaagg 1320
agtcagacca gactgcaga tagtgacctc aaagaagatg gaatttcctc ccggaaatca 1380
agtggcagtg ctgttgagg aaaggaatt gctccagcca gtcccatgct tgggaatgca 1440
agtaatccta ataaggcgga tattcctgaa cgcaagaaaa gctccactgt ccctagtagt 1500
aacacagcat ctggtggaat gacacgacga aatacttatg tttgcagtgaga gagaactaca 1560
gctgatagac actcagtgat tcagaatggc aaagaaaaca gactatttcc tgatcagaga 1620
actccagttg cttcaacaca cagtatcagt agtgacagcca cccagatcg aatccgcttc 1680
ccaagaggca ctgccagtcg tagcactttc cacggccagc cccgggaacg gcgaaccgca 1740
acataataat gccctcctgc ctctcccagc ctgtcccatg aagccacacc attgtcccag 1800
actcgaagcc gaggtccac taatctcttt agtaaatata cttcaaaact cacaaggagg 1860
cttccaactg aatatgagag gaacgggaga tatgagggct caagtcgcaa tgtatctgct 1920
gagcaaaaag atgaaaacaa agaagcaaag cctcgatccc tacgcttcac ctggagcatg 1980
aaaaccacta gttcaatgga tcccggggac atgatgcggg aaatccgcaa agtgttggac 2040
gccataaact gcgactatga gcagaggag cgcttcttgc tcttctgctt ccacggagat 2100
gggcacgcgg agaacctcgt gcagtgaggaa atggaagtgt gcaagctgcc aagactgtct 2160
ctgaacgggg tccggtttta ggggatatcg gggacatcca tagccttcaa aaatattgct 2220
tccaaaattg ccaatgagct aaagctgtaa cccagtgatt atgatgtaaa ttaagtagca 2280
attaaagtgt tttctggaac actgatggaa atgtatagaa taatatttag gcaataacgt 2340
ctgcatcttc taaatcatga aattaaagtc tgaggacgag agcaaaaaaa aaaaaaagg 2400
gcgccctcgc agccgctcga gccgaattcg gctcgaggat tcagtggtg agaggggaaga 2460
aggggaggtt ggggggctcc ttcccttcag aacttgaagt ttctccact gcctcctctc 2520
cagtggtctc ccaggtgcca gacccaaaag ctttctctac agtgataccc ttatattttt 2580
acttcccctt gactcatatg ttttaacatg aattttt 2616

```

<210> 29

<211> 1253

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7478559CB1

<400> 29

```

ctggccctc ctctaccact cccactccct cgccggaccc ccccgccggg gctagcgtct 60
gccgcggtc cgaggggggtg gggctgctgg gaatggctgt gccccttcg gccctcagc 120
cgcgcgctc ctttcacctg aggaggcaca cgcttgccc gcagtgtca tggggcatgg 180
aggagaaggc ggccggccagc gccagctgcc gggagccgcc gggccccccg agggccgccc 240
ccgtcgctg cttcggcatt tccgtggacc cggacgacat ctttcccggg gccctgcgcc 300
tcattccagga gctgcccgcg cattggaaac ccgagcaagt tcggaccaag cgcttcatgg 360
atggcatcac caacaagctg gtggcctgct atgtggagga ggacatgcag gactgcgtgc 420
tggtccgggt gtatggggag cggacggagc tgcgttgga ccgggagaat gaggtcagaa 480
acttccagct gctgcgagca cacagctgtg ccccaaaact ctactgcacc ttccagaatg 540
ggctgtgcta tgagtacatg cagggtgtgg ccctggagcc tgagcacatc cgtgagcccc 600
ggcttttcag gttaatcgcc ttagaaatgg caaagattca tactatccac gccaacggca 660
gcctgcccaa gcccatcctc tggcacaaga tgcacaatta tttcacgctt gtgaagaacg 720
agatcaaccc cagcctttct gcagatgtcc ctaaggtaga ggtgttgga cgggagctgg 780
cctggctgaa ggagcatctg tcccagctgg agtcccctgt ggtgttttgt cacaatgacc 840
tgctctgcaa gaatatcatc tatgacagca tcaaaggcca cgtgcgggtt attgactatg 900
aatatgctgg ctacaactac caagcttttg acattggcaa ccatttcaat gagtttgag 960
gcgtgaatga ggtggattac tgcctgtacc cggcgcgga gaccagctg cagtggtgc 1020
actactacgt gcaggcacia aaggggatgg ccgtgacccc cagggaggtg caaaggctct 1080
acgtgcaagt caacaagttt gccctggcgt ctacttctt ctgggctctc tgggcccctc 1140
tccagaacca gtactccacc atcgactttg atttctcag gtacgcagtg atccgattca 1200

```

accagtactt caaggtgaag cctcaagcgt cagccttggga gatgccaaag tga 1253

<210> 30

<211> 1790

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1698381CB1

<400> 30

tttaagaagg agttccctta taggagatgg aagaaacggc cattaatccg gggacttttt 60  
atgctggaaa caaacctgaa ggtacagggt cggcccgaa gttataccca ccaagagaag 120  
tatgtccgga attgtgggtt ctgcagtcac tgacttcaag aatgaagccg cggaccctcg 180  
cgggtgcagca ttgtactgca agtcaatcga tacaataatt taagtcaact cagctataat 240  
ggaaaagtat gaaaaattag ctaagactgg agaagggtct tatgggggtg tattcaaagt 300  
cagaaacaaa acctctggac aagtagtagc tgtaaaaaa tttgtggaat ctgaagatga 360  
tctgttgtt aagaaaatag cactaagaga aatacgtatg ttgaagcaat taaaacatcc 420  
aaatcttgtg aacctcatcg aggtgttcag gagaaaaagg aaaatgcatt tagtttttga 480  
atactgtgat catacacttt taaatgagct ggaaagaaac ccaatggag ttgctgatgg 540  
agtgatcaaa agcgtattat ggcaaacact tcaagctctt aatttctgtc atatacataa 600  
ctgtattcac agagatataa aacctgaaaa tattctaata actaagcaag gaataatcaa 660  
gatttgtgac ttcgggtttg cacaaattct gattccagga gatgcctaca cggattatgt 720  
agctacgaga tggtaaccgag ctccctgaact tcttgtggga gatactcagt atggttcttc 780  
agtcgatata tgggctattg gttgtgtttt tgcagagctc ctgacaggcc agccactgtg 840  
gcttgaaaaa tcagatgtgg accaacttta tctgataatc agaacactag gaaaattaat 900  
cccaagacat caatcaatct ttaaaagtaa cgggtttttc catggcatca gtatacctga 960  
gccagaagac atggaaactc ttgaggaaaa gttctcagat gttcatcctg tggctctgaa 1020  
cttcattgag ggtgtctctga agatgaatcc agatgacaga ttaacctgtt cccaactcct 1080  
ggagagctcc tactttgatt cttttcaaga ggcccaatt aaaagaaaag cacgtaatga 1140  
aggaagaaac agaagacgcc aacagaatca actgttgccct ctcataccag gaagccacat 1200  
ctccccaca cctgatggaa gaaaacaagt cctccagtta aaatttgatc accttccaaa 1260  
catttaggaa aatgttcttt caagtgc aaa gtaatttaat atgtacacat tttgtacaag 1320  
tgagatagga attctcagtg tttcaaagtc aaatgagcca tatgaaaatt aagatgcctt 1380  
ctagaattgt tttggctctg atcattgctg attcctttcc ccatgctttt acatgccaac 1440  
tttatctttt agaataattt ctttaaagt tataaagcct aaaactgcac atatggaaga 1500  
gacattttca atttcatcag agcagcccct cccgaggcta tctatatgga gaatttgtga 1560  
gcttatactt ggatttatga aaaagattta catgtgtcat cttgcttcag ctgaccacat 1620  
aatttcttaa agcaatatca aatagcctgc ctcaactgtt gtgtaagaaa tgacatatgt 1680  
tctgcatgt gtaattcata cttattgtaa ccagggtctg tgagtattgc tggatatcta 1740  
tactgagtaa atatggtgta gaaagggaac tttgaagggc tgcagattcg 1790

<210> 31

<211> 4132

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474637CB1

<400> 31

ccccgcactg tcttgggtggc agaggggact tttattcagc tggaaccgcg cggcgaggcc 60  
caagtgtctc tggagagatt cgggggttcag gaggtggcgg gtgcacccaa ggggtgctggg 120  
aggaagctcc aggttcccat tcttccccag ggatcggcgt tgccccctgt cgcgggggta 180

gtctagggca	acggaagatg	gcgggcgggc	ccgggacagg	ggttccgggc	tccgctcggg	240
cagagcccac	ccgctgacca	actccgccc	ccccgccc	cgggtgctgt	tccccgagg	300
agtcggagag	gatggcagg	gccggaggcc	agcaccacc	tccgggcgcc	gctggaggag	360
cgccgcggc	agccggcgcc	gcggtcacct	ccgccgctgc	ctcggcgggg	ccgggagagg	420
attcgtctga	cagcgaagcg	gagcaagagg	gaccccagaa	actgatccgc	aaagtgtcta	480
cctcggggca	gatccggacc	aagaccagta	ttaaagaggg	acagctattg	aagcaaacca	540
gttctttcca	aaggtggaaa	aagcgatact	tcaaacttcg	aggccgcacc	ctttactatg	600
caaaggactc	aaagtctctg	atatttgatg	aagttgacct	ctcagatgct	agtgtagctg	660
aagcaagcac	gaaaaatgct	aacaacagct	tcacgatcat	cactccattc	agaaggctaa	720
tgctgtgtgc	tgagaacaga	aaggagatgg	aggattggat	cagctcactg	aagtctgtac	780
agaccagaga	accctacgag	gtggcccagt	ttaatgtgga	acattttctca	gggatgcaca	840
actggtacgc	ctgctcccac	gcccgaacca	ccttctgtaa	cgtgtgcaga	gagagtcttt	900
ctggagtcac	ctcccatggc	ctgtcctgcg	aagtgtgtaa	attcaaggct	cacaaaagat	960
gttcagtgag	agcaacaaat	aactgtaaat	ggactaccct	ggcctccatc	gggaaggaca	1020
ttatagaaga	tgagatggc	gtcgcgatgc	ctcaccagtg	gcttgagggc	aacctgcctg	1080
taagtgccaa	gtgtgctgtc	tgcgacaaaa	catgtggcag	tggtctccgt	ctacaggatt	1140
ggaaatgcct	ttggtgtaag	acaatggtac	acactgcctg	caaagattta	taccatccaa	1200
tatgtccact	tggtcaatgt	aaagtatcta	tcatacctcc	aattgcacta	aacagcaccg	1260
attccgatgg	tttctgtaga	gcaacatttt	cgttctgtgt	tagtcctcta	ttggtttttg	1320
tcaattctaa	gagtggagat	aatcaggagg	taaagttcct	ccgtcgcttt	aaacagttgc	1380
taaateccggc	tcagggtgttt	gatttaatat	atggagggtcc	tcatttaggt	ttaagattat	1440
ttcagaagtt	tgacaatttc	cggattcctg	tatgtggagg	cgatggaagt	gtaggttggg	1500
ttttgtcaga	aatcgataag	ctcaacttga	ataaacagtg	tcagctggga	gtgttgctt	1560
tggttacagg	aatgacctt	gcccaggttc	ttggctgggg	aggttcata	gacgatgaca	1620
cccagcttcc	tcagatccta	gagaaactgg	aacgagccag	tacaaaaatg	ttggacaggt	1680
ggagtataat	gacatatgaa	ctcaaattgc	cacaaaaagc	ttccctactt	ccaggacctc	1740
cagaagcatc	tgaagaattt	tatatgacga	tttatgaaga	ctcagttgca	acgcatctta	1800
caaaaatcct	gaattctgat	gaacatgcag	tggtcatatc	ttctgccaag	acgctatgtg	1860
aaactgtaaa	ggacttcggt	gccaaagtag	aaaagacgta	tgacaaaacc	ttggaaaatg	1920
ccgttgtagc	tgatgcccgt	gccagtaaat	gttcagtcct	aaacgagaag	ctcgaacaac	1980
tgctgcaagg	tttgacacaca	gattcccagg	ctgcgcctgt	tctccctggc	ctcagccctc	2040
tcattgtgga	agaagatgct	gtggaatcgt	ccagtgaaga	gtccctgggt	gaaagcaaag	2100
agcagcttgg	ggatgacgtt	acaaaacctt	cctcccagaa	agccgtcaaa	ccaagggaaa	2160
tcagtgttgc	ggcaaatagt	ttaaagaaag	cagtgaggca	agtcattgag	gaagccggaa	2220
aagttattga	tgaccggaca	gttcaccctc	gtgaaccagc	taatcagtc	tctgattatg	2280
acagcacaga	aacagatgaa	tctaaggagg	aagctaaaga	tgatggtgcc	aaagaatcaa	2340
taactgttaa	aactgcacct	cggctctccag	atgccggggc	aagttatggc	cattcccaaa	2400
ctgattctgt	ccctgggtcca	gctgtggcag	ccagcaaaga	aaacctccct	gtgctcaata	2460
ccagaataat	ctgccaggtt	ttaagagcag	gactggctgc	ctcaattgct	gggagttcga	2520
ttatcaacaa	aatgttactg	gcaaacattg	atccttttgg	tgccacgcgc	tttattgacc	2580
ctgatctaga	ttccgtagat	ggatattcag	aaaaatgtgt	catgaacaat	tactttggga	2640
ttggattaga	tgcaaaaatt	tcattagaat	ttaataataa	aagagaggag	cacctgaaa	2700
aatgcaggag	ccgaactaaa	aacttgatgt	ggtatggagt	ccttggaacc	cgggagtatt	2760
tacagagatc	gtacaagaat	ttagaacaaa	gggttcaact	tgagtgtgat	gggcagtata	2820
ttcctcttcc	cagcttgcaa	ggcatagccg	tggtgaacat	tcccagctat	gctggaggca	2880
ctaacttttg	gggtggaact	aaagaggatg	atataatttgc	tgacacatcc	tttgatgaca	2940
agatcctgga	agttgtagca	atatttgata	gcatgcaa	ggcagtttca	aggggtcatta	3000
aactgcagca	tcacgaata	gccagtgcc	gtacagtga	aatcactata	tttggtgacg	3060
aaggagtccc	agtgaagtg	gatggtgaag	cgtgggttca	gcctccagg	attatcaaaa	3120
ttgtgcacaa	aaacagagca	caaagtctaa	caaggacag	agcctttgag	agcactctga	3180
aatcttggga	agataagcag	aagtgtgatt	ctggtaaacc	agttctccga	acccatttgt	3240
acatccatca	cgccattgac	ttggcaacag	aagaggtgtc	gcagatgcag	ctatgtcccc	3300
aggctgcaga	ggagctcatt	actaggatat	gtgacgcagc	cacaattcac	tgtcttttgg	3360
agcaagaact	ggcccatgct	gtgaatgcct	gctcccatgc	cctgaataaa	gccaacccaa	3420
ggtgcccgga	gagtccttaca	agagacactg	ccactgaaat	agccatcaat	gtgaaggcgc	3480
tgtataatga	aacagaatct	ttgctagttg	gcagggttcc	tttgagctg	gaatcgccac	3540

```

atgaagagcg agtatccaat gccttacact ctgtggaggt ggaattacag aaactgacag 3600
agattccttg gctttattat atcttaacc caaatgaaga tgaggaacct cctatggatt 3660
gcaccaaag gaacaacaga agcacggtat ttcgaatagt gccaaagttt aaaaaggaaa 3720
agggttcagaa gcagaagaca agttcacagc ctggatctgg ggataccgaa agtggggtcat 3780
gtgaagcgaa ttctccaggg aattaaagag cttggaagga gactccaca gtcggaggtg 3840
taatcatatt ggtgctattc cttggaagag aagttattgc cacttaatac aaagtccttg 3900
gaagcaagtg gctgttcttg tagttttctg catagataag taagcaccac tgaagcacct 3960
ctgtggcttg atattttgct gtgggtgaaa ttttgatttg aggtattaga aaatattttt 4020
gtgccgaaca atacattcca cgaagccatt ttctttttgt gcaaacctga catgttcaaa 4080
tatattcaca atggtaataa ggtaggagga atctgagacg attgcattgt ct 4132

```

&lt;210&gt; 32

&lt;211&gt; 1137

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7170260CB1

&lt;400&gt; 32

```

atggaggact ttctgctctc caatgggtac cagctgggca agaccattgg ggaagggacc 60
tactcaaaag tcaaaagaagc attttccaaa aaacacccaaa gaaaagtggc aattaaagtt 120
atagacaaga tgggagggcc agaagagttt atccagagat tctccctcg ggagctccaa 180
atcgtecgta ccctggacca caagaacatc atccagggtg atgagatgct ggagtctgcc 240
gacgggaaaa tctgcctggg gatggagctc gctgagggag gggatgtctt tgactgcgtg 300
ctgaatgggg ggccactgcc tgaaagccgg gccaaagccc tcttcctgca gatggttgag 360
gccatccgct actgccatgg ctgtggtgtg gccaccggg acctcaaatg tgagaacgcc 420
ttgttgacag gcttcaacct gaagctgact gactttggct ttgccaaggt gttgcccag 480
tcacaccggg agctgagcca gaccttctgc ggcagtacag cctatgctgc ccccgagggtg 540
ctgcagggca tccccacga tagcaaaaaa ggtgatgtct ggagcatggg tgtggtcctg 600
tatgtcatgc tctgtgccag cctacctttt gacgacacag acatcccca gatgctgtgg 660
cagcagcaga agggggtgtc cttccccact catctgagca tctcgccga ttgccaggac 720
ctgctcaaga ggctcctgga acccgatatg atcctccggc cttcaattga agaagttagt 780
tggcatccat ggctagcaag cacttgataa aagcaatggc aagtgtcttc caataaagta 840
gggggagaaa gcaaacccaa aaaccgctt ctaaaatggg gatatatatt ttacgcttta 900
agtttactta tcctaaaact tacctacatc taccocagcc ttactactac tctttccttt 960
tagagatctt catggaatca aagggcctca ttcagacttc cttttttttt ttaagagtct 1020
tgctctgtcg ccaggtctgg aatgcagtgg cagattcca gttcactgca actctgcttc 1080
ccaggttcaa gcgattctcc tgcctcagcc tccccagtag ctggctttcc agcacac 1137

```

&lt;210&gt; 33

&lt;211&gt; 3365

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1797506CB1

&lt;400&gt; 33

```

atgagaaggg cggggatcgg cgaggactcc aggtctgggg tgcaggccca gccaggggcg 60
gagccttctc cgggtcgggc ggggacagag cgctcccttg gaggcacca gggacctggc 120
cagccgtgca gctgcccagg cgctatggcg agtgcggtca gggggtcgag gccgtggccc 180
cggctggggc tccagctcca gttcgcgcg ctgctgctcg ggacgtgag tccacaggtt 240
catactctca ggccagagaa cctcctgctg gtgtccacct tggatggaag tctccacgca 300

```

```

ctaagcaagc agacagggga cctgaagtgg actctgaggg atgatcccg catcgaagga 360
ccaatgtacg tcacagaaat ggccctttctc tctgaccag cagatggcag cctgtacatc 420
ttggggaccc aaaaacaaca gggattaatg aaactgccat tcacccatccc tgagctggtt 480
catgcctctc cctgccgcag ctctgatggg gtcttctaca caggccggaa gcaggatgcc 540
tggtttgtgg tggaccctga gtcaggggag acccagatga cactgaccac agaggggtccc 600
tccaccccc gctctacat tggccgaaca cagtatacgg tcacccatgca tgaccaaga 660
gccccagccc tgcgctggaa caccacctac cgcgctact cagcgcctcc catggatggc 720
tcacctggga aatacatgag ccacctggcg tcctgcgga tgggcctgct gctcactgtg 780
gacccaggaa gcgggacggt gctgtggaca caggacctgg gcgtgcctgt gatgggcgtc 840
tacacctggc accaggacgg cctgcgccag ctgccgcac tcacgtggc tcgagacact 900
ctgcatttcc tgcctctccg ctggggccac atccgactgc ctgcctcagg cccccgggac 960
acagccaccc tcttctctac cttggacacc cagctgctaa tgacgctgta tgtggggaag 1020
gatgaaactg gctctatgt ctctaaagca ctggtccaca caggagtggc cctggtgcct 1080
cgtggactga ccctggcccc cgcatgggac cccaccacag atgaggtgac actccaagtc 1140
tcaggagagc gagagggctc acccagcact gctgttagat acccctcagg cagtgtggcc 1200
ctcccaagcc agtggctgct cattggacac cagagctac cccagtcct gcacaccacc 1260
atgctgaggg tccatccac cctggggagt ggaactgcag agacaagacc tccagagaat 1320
acccaggccc cagccttctt cttggagcta ttgagcctga gccgagagaa actttgggac 1380
tccgagctgc atccagaaga aaaaactcca gactcttact tggggctggg accccaagac 1440
ctgctggcag ctagcctcac tgcgtgcctc ctgggagggg ggattctctt tgtgatgagg 1500
cagcagcagg agacccccct ggcaactgca gactttgtc acatctccca ggatgcccg 1560
tccttgcaat cggggggccag ccggaggagc cagaagagge ttcagagtcc ctccactgag 1620
tcaccaccct cctctcccc agctgagcaa ctccagtag tggggaagat ttcctcaat 1680
cccaaggacg tgctgggccc cggggcaggc gggactttcg ttttcagggg acagtgttag 1740
ggacgggcag tggctgtcaa gcggtcctc cgcgagtgtc ttggcctggg tcggcgggaa 1800
gttcaactgc tgcaggagtc tgacaggcac cccaactgct tccgctactt ctgcaccgag 1860
cggggacccc agttccacta cattgccctg gagctctgcc gggcctcctt gcaggagtac 1920
gtagaaaacc cggacctgga tcgccccggg ctggagccc aggtcgtgct gcagcagctg 1980
atgtctggcc tggccacact gcaactctta cacatagtgc accgggacct gaagccagga 2040
aatattctca tcaccgggccc tgacagccag ggcctgggca gagtgggtgt ctcagacttc 2100
ggcctctgca agaagctgcc tgctggccgc tgtagcttca gcctccactc cggcatcccc 2160
ggcagcgaag gctggtatgg ccccgagctt ctgcagctcc tgccaccaga cagtccctacc 2220
agcgtctggg acatcttctc tgacggctgc gctgttctact acgtgcttct tgggtggcagc 2280
cacccttttg gagacagtct ttatcgccag gcaaacatcc tcacaggggc tccctgtctg 2340
gctcacctgg aggaagaggt ccacgacaag gtggttgccc gggacctggg tggagccatg 2400
ttgagccac tgcgcagcc acgccccctc gccccccagg tgctggccca ccccttcttt 2460
tggagcagag ccaagcaact ccagttcttc caggacgtca gtgactggct ggagaaggag 2520
tccgagcagg agccccgtgt gagggcactg gaggcgggag gctgcgcagt ggtccgggac 2580
aactggcacg agcacatctc catgccgctg cagacagatc tgagaaagt ccggtccctat 2640
aaggggacat cagtgcgaga cctgctccgt gctgtgagga acaagaagca ccactacagg 2700
gagctcccag ttgaggtgcg acaggcactc ggccaagtcc ctgatggctt cgtccagtac 2760
ttcacaacac gcttcccacg gctgctcctc cacacgcacc gagccatgag gagctgcgcc 2820
tctgagagcc tcttcttccc ctactaccgc ccagactcag aggcaggag gccatgccc 2880
ggggccacag ggaggtgagg tgggctggat gccacacaga tgggtctcgt gctggctcac 2940
tgaagagctg agcctgtggc tggcctcaga atcaggctgg gtgcagtggc tcacacctgt 3000
aatcccagca ttttgggagg ctgagtgaga ggatcacttg agctcaggag ttcgagacca 3060
gcctggccaa catggcaaca cccatttctt acaaaaaatt tgtaaaatta gccaggcatg 3120
gtggcgcacg cctgtagtcc cagctgcttg ggaggctgag gtgggagaat cacttgagcc 3180
caggagtctg aggtgcaggt gagccaggat catgccactg cactccagcc tggtcacag 3240
agagacactg tcacccccct tccccacaa gactggcaga ggctgggag cctggggctg 3300
atgaagcaga gatgttcgct ggatcccagc tcctggcaca ctgtaaggaa atacaacgaa 3360
gaggt

```

&lt;210&gt; 34

&lt;211&gt; 2049

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1851973CB1

&lt;400&gt; 34

```

gcgttcttttc ccgcggaagt agttgacatt tacaaggagc agcgccccc aaggtcttta 60
gctgttttttt aaggggagaa cagcctttac cctcttttga ctttttcttc gttttttttt 120
tttttggaga cggagtttcg ttctttcgcc caggctggcg tacagtggcg cgatctcggc 180
tcaactgcaac ctctgctccc cgggttcaag cgattctcct gcctcagcct ccggagtagc 240
tggtgattaca ggtgcccgcg accacgcccg gctgatttcc tcttaagact ttctacagct 300
tccttatgaa atcttctgac tgggccttga gcaataaggt cttttgctac aatttagtgc 360
tcttttcttc acactaaatc gaaaactctc cctgttggtc ctgatctgtt tcagtcaggc 420
aaattacatc ctgggaaaac gtcagatgac aggggaggcc actcgcttcc tgctcatcca 480
gtttcgacac tttctgtgct ttcattagct tccagacctc agccctggcc ctgcgtttac 540
tgtacagtca gaactggttt ctacgcctcg cgagggtggg aggtcgtgta tgggaggagg 600
accgcttccc accagcctcg ttgggaagcc aggagaaatc tcttcaaate ctgcgattca 660
gagtcaagtc ccagtcgtcc tttttctggt cggcccagaa ctgtttgtgc ctctccctc 720
atgaggaatg atgtcagtggt ggccgcggtc gccgcccacg aagagtgtaa ggctgcgaag 780
tcggggcttt cccgacgccc cctccgtccg cgtctgcgta ggggaggtga cgagggcggg 840
gcgcgggcgg ggggtgacgt cacggccgcg cgcgcggtgg gcggagcctc actttgaacc 900
cagttggcgg gaatggctgc tcgcgagggt gcagtgtagc cggggccgct gtaggctgtc 960
cagcgatgga tcccaccgcg ggaagcaaga aggagcctgg aggaggcgcg gcgactgagg 1020
agggcggtgaa taggatcgca gtgccaaaac cgccctccat tgaggaattc agcatagtga 1080
agcccattag ccggggcgcc ttcgggaaag tgtatctggg gcagaaaggc ggcaaattgt 1140
atgcagtaaa ggttggttaa aaagcagaca tgatcaacaa aaatatgact catcagggtc 1200
aagctgagag agatgcaact gcaactaagc aaagccatt cattgtccat ttgtattatt 1260
cactgcagtc tgcaacaat gtctacttgg taatggaata tcttattggg ggagatgtca 1320
agtctctcct acatatatat ggttattttg atgaagagat ggctgtgaaa tatatttctg 1380
aagtagcact ggctctagac taccttcaca gacatggaat catccacagg gacttgaaac 1440
cggacaatat gcttatttct aatgagggtc atattaaact gacggatttt ggcttttcaa 1500
aagttacttt gaatagagat attaatatga tggatatcct tacaacacca tcaatggcaa 1560
aacctagaca agattattca agaaccctag gacaagtgtt atcgcttctc agctcgttgg 1620
gatttaacac accaattgca gaaaaaaatc aagaccctgc aaacatcctt tcagcctgtc 1680
tgtctgaaac atcacagctt tctcaaggac tcgtatgcc tatgtctgta gatcaaaagg 1740
acactacgcc ttattctagc aaattactaa aatcatgtct tgaaacagtt gcctccaacc 1800
caggaatgcc tgtgaagtgt ctaacttcta atttactcca gtctaggaaa aggctggcca 1860
catccagtg cagtagtcaa tcccacacct tcatatccag tgtggaatca gaatgccaca 1920
gcagtcceaa atgggaaaaa gattgccagg tttgagggtc atttatctta atgaaaatca 1980
attatgtatg tcaaatgaat gtgagaaata ttatacctt tcatataaat tccataaaga 2040
aatgaaagg                                     2049

```

&lt;210&gt; 35

&lt;211&gt; 2962

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474604CB1

&lt;400&gt; 35

```

accactcgtg ccactgatt atcagcatct ttacttttca ccagcgttct tgggtgtcca 60
cctcctgcgg ccgcggcgga aaacatgacg aaaagcgagg agcagcagcc tctgagtttg 120
caaaaagcct tacagcagtg cgaactgggt caaaacatga tagacttgag catctccaac 180

```

```

ctggaagggc ttaggaccaa atgtgctacc tccaacgacc tcacacaaaa agaaatccgg 240
accctggaga gcaagctggt gaagtacttc agccggcagc tgtcctgcaa aaagaaggta 300
gccttgcaagg agcgcaacgc ggagctggac ggcttcccc agctacggca ctggttccga 360
atcgctcgatg tgcgcaagga ggtcctggag gaaatctccc ccggccagct gagcctggag 420
gacctcttgg agatgacgga tgaacagggtg tgcgagactg tggagaaata cggagccaac 480
cgggaggagt gtgcccgcct caacgcctcc ctctcctgcc tcaggaatgt ccacatgtca 540
ggaggcaacc tttccaaaca agactggacc atccagtggc ccaccacaga gacggggaag 600
gagaacaatc ccgtgtgccc cccggagccc accccgtgga tccgcacca tctctcccag 660
agccccaggg tcccgccaa gtgctgtccag cactattgtc acaccagccc cactcccggg 720
gcccctgtgt acaccacgt ggacaggctt accgtggacg cctacccggg cttgtgcccg 780
ccccgcacc tggagtcggg ccaccgttcc ctgcccccat cgccccgga gcggcacgcg 840
gtccgcaccc cgccgcgcac ccccaacatc gtcaccaccg tgaccccgcc gggcacgccc 900
cccatgagga agaagaacaa gctgaagccc ccggggaccc caccgcctc ctcccgaana 960
ctgatacact tgatcccggg attcaccgag ctgcatcgga gcaaatccca cgagttccag 1020
ctggggcacc gcgtggacga ggcccacag cccaaagcca agaagaagag caaacctctg 1080
aacctcaaga tccacagcag cgtaggcagc tgcgagaaca tcccctctca gcagcgtcc 1140
ccgtctgtgt ccgagcgctc cctccgctcc ttctttgtgg gacacgcacc tttcctgect 1200
tccaccctc ctgttcacac tgaggccaac ttctctgcaa acacactgtc agtgccacgc 1260
tgggtccccg agatccctcg cagagatctt ggcaactcca tcaagcacag gttttccacc 1320
aagtactgga tgtctcagac gtgcacagtc tgtgggaaag ggatgctttt tggcctcaag 1380
tgtaaaaact gcaagttaaa gtgccacaac aaatgcacca aagaagcccc accctgtcat 1440
cttctgatca tccaccgagg agatccagca aggttagtcc ggacagagtc cgttccgtgt 1500
gacatcaaca accctctacg gaagccacct cgctattcag acctgcacat cagtccagacg 1560
ctccccaaaa ccaacaaaat caacaaggac cacatccctg tcccttacca gccagactcc 1620
agcagcaacc cctcctccac gacgtcctcc acgcccctc cgccagcacc cccctcctcc 1680
cctagtgcga cgccgccttc tcccctacac ccttccccac agtgcacacg gcagcagaag 1740
aacttcaacc tgccagcacc ccactactac aaatacaagc agcagttcat cttccagat 1800
gtggtgcccg tgcgggagac gccgacccgg gcgccccagg tcatcctgca tccggtgacc 1860
tcgaatccaa tcttggaagg aaatccatta cttcaaattg aagtggagcc aacgtcggag 1920
aatgaagaagg tccatgatga ggccgaagag tcagaggatg acttcgagga gatgaacctg 1980
tccctcctct cggcccgagg cttcccacgc aaggccagcc agaccagcat cttccttcag 2040
gagtgggaca tcccctttga gcagctggag atcgccgagc tcattggaaa gggccgcttt 2100
gggcaagtgt accacggccg ctggcatggc gaggtggcca tccggctgat tgacattgag 2160
agggacaacg agggaccagt caaggccttc aagcgggagg tgatggccta caggcagaca 2220
cggtgatgaga acgtgggtgt tttcatgggt gcctgcata gccgcctca cctggccatc 2280
atcaccagcc tctgtaaggg acggacgctc tattccgttg tgagggatgc caaatcgctt 2340
ttggatgtca acaaaaccag gcagattgct caagaaattg tgaagggcac gggctacctc 2400
cacgccaagg gaatcctaca caaggacctc aagtc aaaga acgtcttcta tgacaacggc 2460
aaagtggta tcacggactt tggactcttc agcatttctg ggtgctgca ggctggcagg 2520
cgggaggaca aactgcgcac ccagaatggc tggctatgcc acctggcacc agagatcatc 2580
cgccagctgt ccccgacac agaggaggat aagctccctc tctccaagca ctctgacgtc 2640
tttgcccttg gcacaatctg gtatgaactc cacgccaggg aatggccttt caagacccaa 2700
ccagcagagg caataatctg gcaaatgggc acaggcatga aaccaacct cagccagatt 2760
ggcatgggaa aagaaatctc ggacattctt ctcttctgct gggcctttga acaagaagag 2820
agacctacct tcaccaagct catggacatg ctggagaaac tgccaaagcg aaaccgtcgc 2880
ctgtctcacc ctggacattt ctggaagtct gcagagctgt gacctttgga catcgggacg 2940
gcgcccagct gcctgggctc cc 2962

```

&lt;210&gt; 36

&lt;211&gt; 3112

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474721CB1



&lt;400&gt; 36

gggggcattg ctcagcggtg ctaggctggc gcggttgag ccgccgccgg actgacagct 60  
cggtctgcgg accatggaga cctgcgcggg tccacacccg ctgcgcctct tcctctgccg 120  
gatgcagctc tgtctcgcgc tgcttttggg acctggcggg cctgggaccg ccgaggaagt 180  
tatcctcctg gattccaaag cctcccaggc cgagctgggc tggactgcac tgccaagtaa 240  
tgggtgggag gagatcagcg gcgtggatga acacgaccgt cccatccgca cgtaccaagt 300  
gtgcaatgtg ctggagccca accaggacaa ctggctgcag actggctgga taagccgtgg 360  
ccgcgggcag cgcactctcg tggaaactgca gtacacactc cgtgactgca gcagcatccc 420  
tggcgccggc ggtacctgca aggagacctt caacgtctac tacctggaaa ctgaggccga 480  
cctggggcgt gggcgtcccc gcctaggcgg cagccggccc cgcaaaatcg acacgatcgc 540  
ggcggacgag agcttcacgc agggcgacct ggggtgagcgc aagatgaagc tgaacacaga 600  
ggtgcgcgag atcggaccgc tcagccggcg gggtttccac ctggccttcc aggacgtggg 660  
ggcactgcgt gcgcttgtct cgggtgcgcgt ctactacaag cagtgcgcgc ccaccgtgcg 720  
ggccctggcc acgttcccag ccaccgcagc ctgagcgcgc ttctccacac tgggtggaagt 780  
ggccggaacg tgcgtggcgc actcggaaag ggagcctggc agccccccac gcatgcaactg 840  
cgccgcgcag ggcgagtggc tgggtgcctgt gggccgctgc agctgcagcg cgggattcca 900  
ggagcgtggt gacatctgcg aagcctgtcc cccagggttt tacaagggtg ccccgccggc 960  
aagggtctgc tcaccgtgcc cagagcacag ccgggcccctg gaaaacgcct ccaccttctg 1020  
cgtgtgccag gacagctatg cgcgctcacc caccgaccgc ccctcggctt cctgcaccgc 1080  
tgggcccggc tcggcgccgc gggacctgca gtacagcctg agccgctcgc cgtggtgct 1140  
gcgactgcgc tggctgccgc cggccgactc gggaggccgc tcggacgtca cctactcgt 1200  
gctgtgcctg cgtcgcggcc gcgaggggccc ggcggggcgc tcgagccgt gggggccgcg 1260  
cgtggccttc ctaccgcgc aggcagggtc gcgggagcga gccgccacgc tgctgcacct 1320  
gcggcccggg gcgcgtaca cgtgcgcgt ggcgctgctc aacggcgtct cgggcccggc 1380  
ggccgcccgt gttccggttg gcgctgttcc aattaacct ggtagggttg gccctgttcc 1440  
tgttgccggg gttatccgcg accgagtggg accccagagc gtgtccctgt cgtggcggga 1500  
gcccatacct gccggagccc ctggggccaa tgacacggag tacgagatcc gatactacga 1560  
gaagggtcag agtgagcaga cttactccat ggtgaagaca ggggcgccc cagtcaccgt 1620  
caccaacctg aagccggcta cccgctacgt ctttcagatc cgggcccgtt ccccggggccc 1680  
atcctgggag gccagagtt ttaaccccag cattgaagta cagaccctgg gggaggctgc 1740  
ctcagggtcc agggaccaga gccccgccat tgtcgtcacc gtagtgacca tctcgccct 1800  
cctcgtcctg ggtcccgta tgagtgtgct ggccatttgg aggaggccct gcagctatgg 1860  
caaaggagga ggggatgccc atgatgaaga ggagctgtat ttccacttca aagtcccaac 1920  
acgtgcgaca ttccctggacc cccagagctg tggggacctg ctgcaggctg tgcatctgtt 1980  
cgccaaggaa ctggatgcga aaagcgtcac gctggagagg agccttggag gagggcggtt 2040  
tggggagctg tgctgtggct gcttgacgt ccccggtcgc caggagctgc tcgtagccgt 2100  
gcacatgctg agggacagcg cctccgactc acagaggctc ggcttcctgg ccgaggccct 2160  
cacgctgggc cagtttgacc atagccacat cgtgcggctg gagggcgttg ttaccgagg 2220  
aagcaccttg atgattgtca ccgagtacat gagccatggg gccctggacg gcttcctcag 2280  
gcggcacgag gggcagctgg tggctgggca actgatgggg ttgctgcctg ggctggcatc 2340  
agccatgaag tatctgtcag agatgggcta cgttcaccgc ggcctggcag ctgcgccatg 2400  
gctggtcagc agcgacctg tctgcaagat ctctggcttc gggcggggcc cccgggaccg 2460  
atcagaggct gtctacacca ctatgagtgg ccggagccca gcgctatggg ccgctcccga 2520  
gacacttcag ttggccact tcagctctgc cagtgcgtg tggagcttcg gcatcatcat 2580  
gtgggaggtg atggcccttg gggagcggcc ttactgggac atgtctggcc aagacgtgat 2640  
caaggctgtg gaggatggct tccggctgcc accccagg aactgtccta accttctgca 2700  
ccgactaatg ctcgactgct ggcagaagga cccagggtgag cggcccagg tctcccagat 2760  
ccacagcatc ctgagcaaga tgggtgcagga cccagagccc cccaagtgtg ccctgactac 2820  
ctgtcccagg cctcccaccc cactagcgga ccgtgccttc tccaccttc cctccttgg 2880  
ctctgtgggc gcgtggctgg aggcctgga cctgtgccgc tacaaggaca gcttcgcgcc 2940  
tgctggctat gggagccctg aggcctggc cgagatgact gccagaggg acctggtgag 3000  
cctaggcatc tctttggctg aacatcgaga ggcctcctc agcgggatca gcgccctgca 3060  
ggcacgagtg ctccagctgc agggccaggg ggtgcagggt tgagtggacc cc 3112

&lt;210&gt; 37

&lt;211&gt; 3650

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478815CB1

&lt;400&gt; 37

```
caacactcca gagtcgtagg agtgaacact gcacaggaat ctctgcccac ctcaggagaa 60
accaaacttg gggaaaatgt ttgcggtcca cttgatggca ttttacttca gcaagctgaa 120
ggaggaccag atcaagaagg tggacaggtt cctgtatcac atgcggtctc cccgatgacac 180
ccttttggac atcatgaggc ggttcggggc tgagatggag aagggcctgg caaaggacac 240
caaccccacg gctgcagtga agatgttgcc caccttcgtc agggccattc cccgatggttc 300
cgaaaatggg gagttccttt ccctggatct cggaggggtcc aagttccgag tgctgaaggt 360
gcaagtgcgt gaagagggga agcgacacgt gcagatggag agtcagttct acccaacgcc 420
caatgaaatc atccgcggga acggcacaga gctgtttgaa tatgtagctg actgtctggc 480
agatttcatg aagaccaaag atttaaagca taagaaattg ccccttggtc taactttttc 540
tttccctgtg cgacagacta aactggaaga ggggtgtccta ctttcgtgga caaaaaagtt 600
taaggcacga ggagttcagg acacggatgt ggtgagccgt ctgaccaaag ccatgagaag 660
acacaaggac atggacgtgg acatcctggc cctgggtcaat gacaccgtgg ggaccatgat 720
gacctgtgcc tatgacgacc cctactgcca agttggtgtc atcatcggaa ctggcaccaa 780
tgcggtgttac atggaggaca tgagcaacat tgacctggtg gagggcgacg agggcaggat 840
gtgcatcaac acagagtggg gggccttcgg ggacgacggg gccctggagg acattcgcac 900
tgagttcgac agggagctgg acctcggctc tctcaaccca ggaaagcaac tgttcgagaa 960
gatgatcagt ggcctgtacc tgggggagct tgtcaggctt atcttgctga agatggccaa 1020
ggctggcctc ctgtttggtg gtgagaaatc ttctgctctc cacactaagg gcaagatcga 1080
aacacggcac gtggctgcca tggagaagta taaagaaggc cttgctaata caagagagat 1140
cctggtggac ctgggtctgg aaccgtctga ggctgactgc attgccgtcc agcatgtctg 1200
taccatcgct tccttcgctc cggccaatct ctgtgcagca gctctggcgg ccatcctgac 1260
acgcctcagg gagaacaaga aggtggaacg gctccggacc acagtgggca tggacggcac 1320
cctctacaag atacaccctc agtacccaaa acgcctgcac aaggtggtga ggaaactggt 1380
cccaagctgt gatgtccgct tcctcctgtc agagagtggc agcaccaagg gggccgccat 1440
ggtgaccgag gtggcctccc gcgtgcaggc ccagcggaaag cagatcgaca ggggtgctggc 1500
ttgttccag ctgaccgagc agcagctcgt ggacgtgcag gccaaagatgc gggctgagct 1560
ggagtatggg ctgaagaaga agagccacgg gctggccacg gtcaggatgc tgcccaccta 1620
cgtctgcggg ctgccggacg gcacagagaa aggaaagttt ctgccctgg atcttggggg 1680
aaccaacttc cgggtcctcc tgggtgaagat cagaagtggc cggaggtcag tgcgaatgta 1740
caacaagatc ttcgccatcc ccttgagat catgcagggc actggtgagg agctctttga 1800
tcacattgtg cagtgcacgc ccgacttcct ggactacatg ggcctcaagg gaggcctcct 1860
acctttgggc ttcacattct catttcctct caggcagatg agcattgaca agggaaact 1920
catagggtgg accaaaggtt tcaaggccac tgactgtgaa ggggaggacg tgggtggacat 1980
gctcagggaa gccatcaaga ggagaaacga gtttgacctg gacattgttg cagtcgtgaa 2040
tgatacagtg gggaccatga tgacctgtgg ctatgaagat cctaattgtg agattggcct 2100
gattgcagga acaggcagca acatgtgcta catggaggac atgaggaaca tcgagatggt 2160
ggaggggggt gaagggaaga tgtgcatcaa tacagagtgg ggaggatttg gagacaatgg 2220
ctgcatagat gacatccgga cccgatacga cacggagggt gatgaggggt ccttgaatcc 2280
tggcaagcag agatacgaga aaatgaccag tgggatgtac ttgggggaga ttgtgcggca 2340
gatcctgacg gacctgacca agcagggtct cctcttccga gggcagattt cagagcgtct 2400
ccggaccagg ggcactctcg aaaccaagtt cctgtcccag atcgaaagcg atcggctggc 2460
ccttctccag gtcaggagga ttctgcagca gctgggcctg gacagcacgt gtgaggacag 2520
catcgtggtg aaggagggtg gcgagccgt gtcccggcgg gcggcccagc tctgcggtgc 2580
tggcctggcc gctatagtgg aaaaaaggag agaagaccag gggctagagc acctgaggat 2640
cactgtgggt gtggacggca ccctgtacaa gctgcaccct cacttttcta gaattattga 2700
ggaaactgtg aagggaactag cccctcgatg tgatgtgaca ttcattgtgt cagaagatgg 2760
cagtggaaaa ggggcagcac tgatcactgc tgtggccaag aggttacagc aggcacagaa 2820
ggagaactag gaacccctgg gattggacct gatgcatctt ggatactgaa cagcttttcc 2880
```

```

tctggcagat cagttgggtca gagaccaatg ggcaccctcc tggctgacct caccttctgg 2940
atggccgaaa gagaacccca ggttctcggg tactcttagt atcttgtaet ggatttgcag 3000
tgacattaca tgacatctct atttgggtata tttgggccaa aatgggccaa cttatgaaat 3060
caaagtgtct gtcctgagag atcccccttc aacacattgt tcagggtgagg cttgagctgt 3120
caattctcta tggctttcag tcttgtggct gcgggacttg gaaatatata gaatctgcc 3180
atgtggctgg caggctgttt ccccatggg atgcttaagc catctcttat aggggattgg 3240
accctgtact tgtggatgaa cattggagag caagaggaaac tcacgttatg aactaggggg 3300
atctcatcta acttgcctt aacttgccat gttgacttca aacctgttaa gagaacaaag 3360
actttgaagt atccagcccc aggggtgcaga gaggttgatt gccaggggagc actgcaggaa 3420
tcattgcatg cttaaagcga gttatgtcag caccctgtag gattttgttc cttattaagt 3480
gtgtgccatg tgggtgggtg ctgtctgggg catctgtttt tcattttgcc tgtgggtttgt 3540
gttgccagggt ttgatagttg ttttaaggat tgtaggtat aggaaatcca gtaaattaat 3600
aaaaaaattt tgattttcca ataaaaaaaa aaaaaaaaaa aaaaaaaaaa 3650

```

&lt;210&gt; 38

&lt;211&gt; 7789

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477141CB1

&lt;400&gt; 38

```

cacaccctga aagccggtcc ctggccgtgc tggcccccct gcaggacgtg gacgtggggg 60
ccggggagat ggcgtgttt gagtgccgtg tggccggggc cactgacgtg gaggtggatt 120
ggctgtgccg tggccgcctg ctgcagcctg cactgctcaa atgcaagatg catttcgatg 180
gccgcaaagt caagctgcta cttacatctg tacatgagga cgacagtggc gtctacacct 240
gcaagctcag cacggccaaa gatgagctga cctgcagtgc ccggctgacc gtgcggccct 300
cgttggcacc cctgttcaca cggctgctgg aagatgtgga ggtgttgag ggccgagctg 360
cccggttctga ctgcaagatc agtggcaccc cgccccctgt tgttacctgg actcattttg 420
gctgccccat ggaggagagt gagaacttgc ggctgcccga ggacgggggt ctgcactcac 480
tgcacattgc ccatgtgggc agcgaggacg aggggctcta tgcggtcagt gctgttaaca 540
cccatattgc ggccactgc tcagcccagc tgtagtaga agagccccgg acagccgcct 600
caggccccag ctggaagctg gagaagatgc catccattcc cgaggagcca gagcagggtg 660
agctggagcg gctgtccatt cccgacttcc tgcggccact gcaggacctg gaggtgggac 720
tggccaagga ggccatgcta gagtgccagg tgaccggcct gccctacccc accatcagct 780
ggttccacaa tggccaccgc atccagagca gcgacgaccg gcgcatgaca cagtacaggg 840
atgtccatcg cttggtgttc cctgccgtgg ggctcagca cgccggtgtc tacaagagcg 900
tcattgccaa caagctgggc aaagctgcct gctatgccca cctgtatgtc acagatgtgg 960
tcccaggccc tccagatggc gccccgcagg tgggtggctgt gacggggagg atggtcacac 1020
tcacatggaa cccccccagg agtctggaca tggccatcga cccggactcc ctgacgtaca 1080
cagtgcagca ccagggtgct ggctcggacc agtggacggc actggtcaca ggccctgcgg 1140
agccagggtg ggccagccaca gggctgcgta aggggggtcca gcacatcttc cgggtcctca 1200
gcaccactgt caagagcagc agcaagccct caccaccttc tgagcctgtg cagctgctgg 1260
agcacggccc aacctgggag gaggccccct ccatgtgtga caaaccagac atcgtgtatg 1320
tgggtggaggg acagcctgcc agcgtcaccg tcacattcaa ccatgtggag gccagggtcg 1380
tctggaggag ctgccagggg gccctcctag aggcacgggc cgggtgtgtac gagctgagcc 1440
agccagatga tgaccagtac tgtcttcgga tctgccgggt gagccgccgg gacatggggg 1500
ccctcacctg caccgcccga aaccgtcacg gcacacagac ctgctcggtc acattggagc 1560
tggcagaggc ccctcggttt gagtccatca tggaggacgt ggagggtggg gctggggaaa 1620
ctgctcgctt tgcggtgggt gtcgagggaa aaccactgcc ggacatcatg tggtaaaagg 1680
acgagggtgt gctgaccgag agcagccatg tgagcttcgt gtacgaggag aatgagtgtc 1740
ccctgggtgt gctcagcacg ggggccagg atggaggcgt ctacacctgc accgccaga 1800
acctggcggg tgagggtctc tgcaaagcag agttggctgt gcattcagct cagacagcta 1860
tggaggtcga gggggtcggg gaggatgagg accatcgagg aaggagactc agcgactttt 1920

```

atgacatcca ccaggagatc ggcaggggtg ctttctccta cttgaggcgc atagtggagc 1980  
 gtagctccgg cctggagttt ggggccaaagt tcatccccag ccaggccaag ccaaaggcat 2040  
 cagcgcgctcg ggaggcccggt ctgctggcca ggctccagca cgactgtgtc ctctacttcc 2100  
 atgaggcctt cgagaggcgc cggggactgg tcattgtcac cgagctctgc acagaggagc 2160  
 tgctggagcg aatcgccagg aaaccaccg tggtgtagtc tgagatccgg gcctatatgc 2220  
 ggcaggtgct agagggaata cactacctgc accagagcca cgtgctgcac ctcgatgtca 2280  
 agcctgagaa cctgctgggtg tgggatgggtg ctgaggggcga gcagcaggtg cggatctgtg 2340  
 actttgggaa tgcccaggag ctgactccag gagagcccca gtactgccag tatggcacac 2400  
 ctgagtttgt agcaccgag attgtcaatc agagccccgt gtctggagtc actgacatct 2460  
 ggctgtggg tgttgttgc ttcctctgtc tgacaggaat ctccccgtt gttggggaaa 2520  
 atgaccggac aacattgat aacatccgaa actacaacgt ggccttcgag gagaccacat 2580  
 tcctgagcct gagcaggag gcccggggct tcctcatcaa agtgttggtg caggaccggc 2640  
 tgagacctac cgcagaagag accctagaac atccttggtt caaaactcag gcaaagggc 2700  
 cagaggtgag cacggatcac ctgaagctat tcctctcccg gcggaggtg cagcgctccc 2760  
 agatcagcta caaatgccac ctggtgctgc gccccatccc cgagctgtg cgggcccccc 2820  
 cagagcgggt gtgggtgacc atgccagaa ggccaccccc cagtgggggg ctctcatcct 2880  
 cctcggtatc tgaagaggaa gagctggaag agctgcctc agtgccccgc cactgcagc 2940  
 ccgagttctc tggctcccg gtgtccctca cagacattcc cactgaggat gaggccctgg 3000  
 ggaccccaga gactgggggt gccaccccca tggactggca ggagcaggga agggctccct 3060  
 ctgaggacca ggaggtccc agcccagagg ccctccctc cccaggccag gagcccgag 3120  
 ctggggctag cccagggcg ggagagctcc gcaggggag ctgggtgag agcgccctgc 3180  
 cccgggcccg gccggggag ctgggcccgg gcctgcacaa ggcgcgctc gtggagctgc 3240  
 cgcagcgccg gagccccgc cggggagcca cccgctggc cggggaggc ctgggtgagg 3300  
 gcgagtatgc ccagaggctg caggccctgc gccaggggt gctgcgggga ggccccgag 3360  
 atggcaaggt cagcggcctc aggggtcccc tgctggagag cctggggggc cgtgctcggg 3420  
 acccccggtt ggcacgagct gctccagcg aggcagcgcc ccaccaccag cccccactc 3480  
 agaaccgggg cctgcaaaag agcagcagct tctcccagg tgaggcggag ccccggggccc 3540  
 ggcaccgccc agcggggggc cccctcgaga tccccgtggc caggcttggg gcccgtaggc 3600  
 tacaggagtc tccttccctg tctgccctca gcgaggccca gccatccagc cctgcacggc 3660  
 ccagcgccgc caaaccaggt acccctaagt ctgcagaacc ttctgccacc acacctagt 3720  
 atgctcgca gcccccgca cccagcctg cccaagacaa ggctccagag cccaggccag 3780  
 aacctcctc agcttccaag cctgcaccac cccccaggc cctgcaaac ctacgctgc 3840  
 cctcacctac ctatctcag atcattcag cctccagct gtcaggccac gccaggggc 3900  
 cctcgagggt cctgcccgc ccgcttcag agcccaagcc ccacgctgt gtctttgca 3960  
 ggggtggctc cccacctcg ggagcccccg agaagcgct gccctcagcc gggggtcccc 4020  
 cgggtgctag cgagaaagcc cgagtccca cgggtgcccc caggccaggc agcagtctca 4080  
 gtagcagcat cgaaaacttg gagtccgagg ccgtgttcca ggccaagtcc aagcgagcc 4140  
 gcgagtcgcc cctgtcgctg gggctgcggc tgctgagccg ttccgctcg gaggagcgc 4200  
 gccccttcgg tggggccgag gaggaggatg gcatataccg gccagcccg gcggggaccc 4260  
 cgctggagct ggtgcgacgg cctgagcgct cacgctcggg gcaggacctc agggctgtcg 4320  
 gagagcctgg cctcgtccgc cgcctctcgc tgtactgtc ccagcggtc cggcggaccc 4380  
 ctcccgcgca gcgccaccg gcctgggagg ccgcggcg ggacggagag agctcggagg 4440  
 gcgggagctc ggcgcggggc tccccgtgc tggcgatgc caggcggctg agcttcacc 4500  
 tggagcggtc gtccagccga ttgcagcgca gtggcagcag cgaggactcg gggggcgct 4560  
 cgggcccag cagcgcgctg ttccgacggc ttccgagggc cacgtccag ggcgagagtc 4620  
 tgcggcgctc tggccttcgc cacaaccagt tggccgcccc ggccggcgcc accacgcctt 4680  
 ccgcccagtc cctgggctcc gaggccagcg ccacgtcggg ctctcagcc ccaggggaaa 4740  
 gccgaagccg gctccgctgg ggcttctctc ggcccgggaa ggacaagggg ttatcgccac 4800  
 caaacctctc tgccagcgtc caggaggagt tgggtacca gtacgtgcgc agtgagtcag 4860  
 acttcccccc agtcttccac atcaaaactca aggaccaggt gctgctggag ggggaggcag 4920  
 ccacctgct ctgctgcca gcggcctgcc ctgcaccgca catctcctgg atgaaagaca 4980  
 agaagtccct gaggtcagag ccctcagtga tcactgtgtc ctgcaaagat gggcggcagc 5040  
 tgctcagcat ccccgggcg ggcaagcggc acgcccgtct ctatgagtgc tcggccacca 5100  
 acgtactggg cagcatcac agctcctgta ccgtggctgt gggccgagtc ccaggaaagc 5160  
 tagctcctcc agaggtaccc cagacctacc aggcacggc gctggtgctg tggaaagccg 5220  
 gagacagccg ggcacctgac acgtatacgc tggagcgggc agtggtggg gagtctgtgt 5280

```

ggcaccctgt gagctcaggc atccccgact gttactacaa cgtgaccac ctgccagttg 5340
gcgtgactgt gaggttcctg gtggcctgtg ccaaccgtgc tgggcagggg cccttcagca 5400
actcttctga gaaggctctt gtcagggtga ctcaagattc ttcagctgtg ccactctgtg 5460
cccaccaaga ggccccctgt acctcaaggc cagccagggc cgggcctcct gactctccta 5520
cctcactggc cccaccccta gctcctgtgt cccccacac cccgtcagtc actgtcagcc 5580
cctcatctcc cccacacact cctagccagg ccttgtcttc gctcaaggct gtgggtccac 5640
caccaccaac ccctccacga agacacaggg gcctgcaggc tgcccgcca ggggagcca 5700
ccctaccag taccacgtc accccaagt agccaagcc ttctgtcctt gacactggga 5760
ccccgatccc agcctccact cctcaagggt taaaccagt gtcttcctct actcctgtgt 5820
atgtgggtgac ttctttgtg tctgcaccac cagcccctga gccccagcc cctgagcccc 5880
ctcctgagcc taccaagggt actgtgcaga gcctcagccc ggccaaggag gtgggtcagct 5940
cccctgggag cagtccccga agctctccca ggctgagggt taccactctt cgacagggtc 6000
cccctcagaa accctacacc ttctggagg agaaagccag gggccgcttt ggtgttgtgc 6060
gagcgtgccg ggagaatgcc acggggcgaa cgttcgtggc caagatcgtg ccctatgtg 6120
ccgagggcaa gcggcggtgc ctgcagggt gctgaggtgt gcggaacctg caccacgagc 6180
ggatcatgtc cctgcacgag gcctacatca cccctcggta cctcgtgtgc attgtgaga 6240
gctgtggcaa ccgggaactc ctctgtgggc tcagtgcagc gttccggtat tctgaggatg 6300
acgtggccac ttacatggtg cagctgtctc aaggcctgga ctacctccac ggccaccacg 6360
tgctccacct agacatcaag ccagacaacc tgctgtggc ccctgacaat gccctcaaga 6420
ttgtggactt tggcagtgcc cagccctaca acccccaggc ccttaggccc cttggccacc 6480
gcacgggcac gctggagtgc atggctccgg agatggtgaa gggagaacct atcggctctg 6540
ccacggacat ctggggagcg ggtgtgctca cttacattat gctcagtgga cgctccccgt 6600
tctatgagcc agacccccag gaaacggagg ctccgattgt ggggggccgc tttgatgct 6660
tccagctgta cccaataca tcccagagcg ccacctctt ctgcaaaag gttctctctg 6720
tacctcctg gagccggccc tccctgcagg actgcctggc ccacctatgg ttgcaggagc 6780
cctacctgat gaagctgcgc cgcagacgc tcacctcac caccaaccgg ctcaaggagt 6840
tccctggcga gcagcgccgg cgcgggctg aggtgccac ccgccacaag gtgctgtgc 6900
gctcctaccc tggcgccccc tagaggcacg gaccacagcc aggcctcggg cttcaactgg 6960
ggttcccacc aatgccacgg gacattccag ggccacgct gagccaggcg ggcctggggc 7020
ttcggttacc accagcagca acatctggct gggctcttac ctcatagacc ttcaaggaca 7080
gagacccag ggctggacc tgatgccacc ccaggccaaa gccagagtgg gagaccatt 7140
ggtcaggctc agcagggtgg gaacaggcag agggacaaga ggggaatgga gaagtggaga 7200
ggaaaaggaa tccagggaca ggaaggggga ggctctagga aggttctggg ttgggggtca 7260
gtccatctca gggagaacca aggaagggtg gcatggctgg agaggaggaa aaggaaggag 7320
cctcagggtg cagggcagta ggctgggagt cagtgtggca aagcgggggc aggacacaga 7380
tacagtggca ggggcccagg gctgggacat gagagaaggc agcgaaggcg cagagggaga 7440
agagaggact cagggtggagg tggggtgggt cagctgtcag catccctcag aggagaaatg 7500
tggagagctg gaggccagca gtcactcaca ctctgtctgt cctcctgtcc agtggataca 7560
gccctgggcg ctctgtctgg ccaaggatgt cccactgcc cctccatggc ctttggcctt 7620
cttccattc atatttattt atttattgac ttttatgaag tttcccttc catccgatcc 7680
ctactgccca tgttgtctg accatccctc ccagccatcc agctgtctgt ctgtctgcca 7740
caaggaaata aaaatggcaa gcagcataaa aaaaaaaaaa aaaaaaaaaa 7789

```

&lt;210&gt; 39

&lt;211&gt; 1937

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2190612CB1

&lt;400&gt; 39

```

gtggtgtggc tgcagtggag agttcccaac aaggctacgc agaagaacct ccttgactga 60
agcaatggag gggggtccag ctgtctgtgt ccaggatcct cgggcagagc tggtagaacg 120
ggtggcagcc atcgatgtga ctcacttgga ggaggcagat ggtggcccag agcctactag 180

```

```

aaacggtgtg gacccccac cacggggccag agctgcctct gtgatccctg gcagtacttc 240
aagactgtct ccagcccgcc ctagcctctc agccaggaag ctttccctac aggagcggcc 300
agcaggaagc tatctggagg cgcaggctgg gccttatgcc acggggcctg ccagccacat 360
ctcccccggg gcctggcgga ggccaccat cgagtccac cacgtggcca tctcagatgc 420
agaggactgc gtgcagctga accagtacaa gctgcagagt gagattggca aggggtgccta 480
cgggtgtggtg aggctggcct acaacgaaag tgaagacaga cactatgcaa tgaaagtcct 540
ttccaaaaag aagttactga agcagtatgg ctttccacgt cgccctcccc cgagagggtc 600
ccaggctgcc cagggaggac cagccaagca gctgctgccc ctggagcggg tgtaccagga 660
gattgccatc ctgaagaagc tggaccacgt gaatgtggtc aaactgatcg aggtcctgga 720
tgaccagctc gaggacaacc tctatttggg gtttgacctc ctgagaaagg ggcccgtcat 780
ggaagtgcce tgtgacaagc ccttctcgga ggagcaagct cgcctctacc tgcgggacgt 840
catcctgggc ctcgagtact tgcactgcca gaagatcgtc cacagggaca tcaagccatc 900
caacctgtct ctaggggatg atgggcacgt gaagatcgcc gactttggcg tcagcaacca 960
gtttgagggg aacgacgctc agctgtccag cacggcggga accccagcat tcatggcccc 1020
cgaggccatt tctgattccg gccagagctt cagtgggaag gccttggatg tatgggccac 1080
tggcgtcacg ttgtactgct ttgtctatgg gaagtgcctg ttcacgacg atttcatcct 1140
ggccctccac aggaagatca agaattgagc cgtggtgttt cctgaggagc cagaaatcag 1200
cgaggagctc aaggacctga tcctgaagat gttagacaag aatcccgaga cgagaattgg 1260
ggtgccagac atcaagttgc acccttgggt gaccaagaac ggggaggagc cccttctctc 1320
ggaggaggag cactgcagcg tgggtggaggt gacagaggag gaggttaaga actcagtcag 1380
gtcatcccc agctggacca cgggtgatcct ggtgaagtcc atgctgagga agcgttcctt 1440
tgggaacccg tttgagcccc aagcacggag ggaagagcga tccatgtctg ctccaggaaa 1500
cctactggtg aaagaagggt ttggtgaagg gggcaagagc ccagagctcc ccggcgctcca 1560
ggaagacgag gctgcacct gagccctgc atgcacccag ggccacccgg cagcacactc 1620
atcccgccgc tccagaggcc caccctcat gcaacagccg ccccgaggg cagggggctg 1680
gggactgcag cccactccc gcccctcccc catcgctgctg catgacctcc acgcacgcac 1740
gtccagggac agactggaat gtatgtcatt tggggtcttg ggggcagggc tcccacgagg 1800
ccatcctcct cttcttggac ctcttggcc tgagccattc tgtggggaaa ccgggtgccc 1860
atggagcctc agaaatgaca cccggctggt tggcatggcc tggggcagga ggcagaggca 1920
ggagaccaag atggcag

```

1937

&lt;210&gt; 40

&lt;211&gt; 5373

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477549CB1

&lt;400&gt; 40

```

atggagcggc ggctgcgcgc gctggagcag ctggcgcggg gcgaggccgg cggtgcccc 60
gggctcgacg gcctcctaga tctgctgctg gcgctgcacc acgagctcag cagcgcccc 120
ctacggcggg agcgcgagct ggcgagttc ctgagctggg ccagcccctt cgtatcaaag 180
gtgaaagaac tgcgtctgca gagagatgac tttgagatct tgaaggatgat cggccgagga 240
gcctttgggg aggtcaccgt ggtgaggcag agggacactg ggagattttt tgccatgaaa 300
atgctgcaca agtgggagat gctgaagagg gctgagacag cctgtttctg ggaggagcgg 360
gatgtgctcg tgaaagggga cagccgttgg gtgaccactc tgcactatgc cttccaagac 420
gaggagtacc tgtaccttgt gatggactac tatgctggtg gggacctcct gacgctgctg 480
agccgcttgc aggaccgtct cccgcccag ctggcccagt tctacctggc tgagatgggtg 540
ctggccatcc actcgctgca ccagctgggt tatgtccaca gggatgtcaa gccagacaac 600
gtcctgctgg atgtgaacgg gcacattcgc ctggctgact tcggctcctg cctgcgtctc 660
aacaccaacg gcatggtgga ttcatactg gcagtaggga cgccggacta tatctcccct 720
gagatcctgc aggccatgga ggagggcaag ggccactacg gccacagtg tgactggtgg 780
tcgcttgagg tctgcgccta tgagctgctc tttggggaga cgcccttcta tgctgagtcc 840
ttggtggaaa cctacggcaa gatcatgaac cacaggacc acctgcagtt ccccccggac 900

```

gtgcctgacg	tgccagccag	cgcccaagac	ctgatccgcc	agctgctgtg	tgcgccaggaa	960
gagcggctag	gccgtggtgg	gctggatgac	ttccggaacc	atcctttctt	cgaaggcgtg	1020
gactgggagc	ggctggcgag	cagcacggcc	ccctatatcc	ctgagctgcg	ggggcccatg	1080
gacacctcca	actttgatgt	ggatgacgac	accctcaacc	atccaggggac	cctgccaccg	1140
ccctcccacg	gggccttctc	cggccatcac	ctgccattcg	tgggcttcac	ctacacctca	1200
ggcagtcaca	gtcctgagag	cagctctgag	gcttgggctg	ccctggagcg	gaagctccag	1260
tgtctggagc	aggagaaggt	ggagctgagc	aggaagcacc	aagaggccct	gcacgcccc	1320
acagaccatc	gggagctgga	gcagctacgg	aaggaagtgc	agactctgcg	ggacaggctg	1380
ccagagatgc	tgagggacaa	ggcctcattg	tcccagacgg	atgggcccc	agctggtagc	1440
ccaggtcagg	acagtgcact	acggcaggag	cttgaccgac	ttcaccggga	gctggccgag	1500
ggtcgggagc	ggctgcaggc	tcaggagcag	gagctctgca	ggggccaggg	gcagcaggag	1560
gagctgcttc	agaggctaca	ggaggcccag	gagagagagg	cggccacagc	tagccagacc	1620
cgggcccctg	gtccccagct	ggaggaagcc	cgggctgccc	agagggagct	ggaggcccag	1680
gtgtcctccc	tgcgcgggca	ggtgacgcag	ctgcaggggac	agtgggagca	acgccttgag	1740
gagtcgtccc	aggccaagac	catccacaca	gcctctgaga	ccaacgggat	gggaccccc	1800
gaggggtggg	ctcaggaggc	ccaactgagg	aaggagggtg	ccgccctgcg	agagcagctg	1860
gagcaggccc	acagccacag	gccgagtggg	aaggaggagg	ctctgtgcca	gctgcaggag	1920
gaaaaccgga	ggctgagccg	ggagcaggag	cggctagaag	cagagctggc	ccaggagcag	1980
gagagcaagc	agcggctgga	gggtgagcgg	cgggagacgg	agagcaactg	ggaggcccag	2040
ctcgccgaca	tcctcagctg	ggtgaatgat	gaggaaggtc	caagaggcta	cctgcaggcc	2100
ctggccacca	agatggcaga	ggagctggag	tccttgagga	acgtaggcac	ccagacgctc	2160
cctgccccgg	cactgaagat	ggaggcctcg	gccaggctgg	agctgcagtc	agcgtgggag	2220
gccgagatcc	gcgccaagca	gggcctgtag	gagcggctga	cacaggtgca	ggaggcccag	2280
ctgcaggctg	agcggctct	gcaggaggcc	gagaagcaga	gccaggccct	gcaacaggag	2340
ctcgccatgc	tgcgggagga	gctgcggggc	cgaaggccag	tggacaccaa	gccctcaaac	2400
tccttgattc	ccttcctgtc	cttcgggagc	tcagagaagg	attctgcca	ggaccctggc	2460
atctcaggag	aggccacaag	gcatggagga	gagccagatc	tgaggccgga	gggccgacgc	2520
agcctgcgca	tgggggctgt	gttccccaga	gcacccactg	ccaacacagc	ctctacagaa	2580
ggtcttcctg	ctaagggatg	gggcatgggg	ccctgggagg	ccttgggtaa	tggctgtccc	2640
cctccccagg	ccggctcaca	cacgctgcgc	ccccggagct	tcccatcccc	gaccaagtgt	2700
ctccgctgca	cctcgctgat	gctgggcctg	ggccgccagg	gcctgggttg	tgatgcctgc	2760
ggctactttt	gtcacacaac	ctgtgcccc	caggccccac	cctgccccgt	gccccctgac	2820
ctctctcgca	cagccctggg	agtacacccc	gaaacaggca	caggcactgc	ctatgagggc	2880
ttctctcggt	tgcgcgggcc	ctcagggtgc	cggcggggct	ggcagcgctg	gtttgtgcc	2940
ctgagtgact	cacgcctgct	gctgtttgac	gcccctgacc	tgaggctcag	cccgccagct	3000
ggggccctcc	tgcaggtcct	agatctgagg	gacccccagt	tctcggttac	ccctgtcctg	3060
gcctctgatg	ttatccatgc	ccaatccagg	gacctgccac	gcatctttag	ggtgacaacc	3120
tcccagctgg	cagtgcgcgc	caccacgtgc	actgtgctgc	tgctggcaga	gagcgagggg	3180
gagcgggaac	gctggctgca	ggtgctgggt	gagctgcagc	ggctgctgct	ggacgcgcgg	3240
ccaagacccc	ggcccgtgta	cacactcaag	gaggcttacg	acaacgggct	gccgctgctg	3300
cctcacacgc	tctgcgtgct	catectcgac	caggatcgac	ttgcgcttgg	caccgaggag	3360
gggctctttg	tcacccatct	gcgcagcaac	gacatcttcc	aggtggggga	gtgccggcgc	3420
gtgcagcagc	tgaccttgag	ccccagtgca	ggcctgctgg	tcgtgctgtg	tggccgcggc	3480
cccagcgtgc	gtctctttgc	cctggcggag	ctggagaaca	tagaggtagc	aggtgccaa	3540
atccccaggt	ctcagggtctg	ccagggtgctg	gcagctggaa	gcatcctgca	ggcccgccac	3600
ccggtgctct	gtgtagccgt	caagcgccag	gtgctctgct	accagctggg	cccgggccct	3660
gggcccctgg	agcggccgat	ccgtgagctg	caggcacctg	ccactgtgca	gagcctgggg	3720
ctgctgggag	accggctatg	tgtgggcgcc	gccgggtggc	ttgactcta	cccgtgctc	3780
aacgaggctg	cgccgttggc	gctggggggc	ggtttgggtg	ctgaggagct	gccaccatcc	3840
cgcggggggc	tgggtgaggc	actgggtgcc	gtggagctta	gcctcagcga	gttctctgta	3900
ctcttcacca	ctgctggcat	ctacgtggat	ggcgcaggcc	gcaagtctcg	tggccacgag	3960
ctgttggtgg	cagcagcgcc	catgggctgg	gggtatggcg	ccccctacct	gacagtgttc	4020
agcgagaact	ccatcgatgt	gtttgacgtg	aggagggcag	aatgggtgca	gaccgtgccg	4080
ctcaagaagg	tgcggcccct	caatccagag	ggctccctgt	tcctctacgg	caccgagaag	4140
gtccgcctga	ctcacctcag	gaaccagctg	gcagagaagg	acgagttcga	catcccggac	4200
ctcaccgaca	acagccggcg	ccagctgttc	cgcaccaaga	gcaagcgccg	cttctttttc	4260

cgcggtgtcgg aggagcagca gaagcagcag cgcagggaga tgctgaagga cccttttgtg 4320  
cgctccaagc tcatctcgcc gcctaccaac ttcaaccacc tagtacacgt gggccctgcc 4380  
aacgggcgcc ccggcgccag ggacaagtcc ccgtcccagc ccctccgcac tgtcacccaa 4440  
caggctcccg aagagaaggg ccgagttgcc cgcggctccg gccacacagc gcccacagc 4500  
ttctccgagg cgttcggcg cccagcctcc atgggcagcg aaggcctcgg tggagacgca 4560  
gacccactg gagcagtga gaggaaccc tggacatccc tgtccagcga gtctgtgtcc 4620  
tgcccccagg gatcgctgag ccctgcaacc tccctaattgc aggtctcaga acggccccga 4680  
agcctcccc tgtcccctga attggagagc tctccttgat gccctctgtt agggcccacc 4740  
ccaatcccag ggcagaagga catgaggagc caaagagctt gaggaatgcc atactccggc 4800  
tggtcggga catggaaatt cggactcagg gaggaccgg gctgggcaat gactgggaga 4860  
cttgctggg ttcccaggac ttgggggtcc tgactcccag ccctcatcct gccttaaccc 4920  
tctgttccca gcccagcct ttctaagcca ttgggaatag aatggcccct tttgttctgg 4980  
tgtccaggg tgattgtgcc aaagctctta ttccagtgcc caagccccca gaggcttgta 5040  
agagtggga tgagggatgg agagggactg ggtctctggg aacagggttg aggtcttacc 5100  
tgtggactgt ctgactccca gctgaggcca agatggggca tgtcccgcgc tctgcttagc 5160  
gtctgggtga gaaaaacagg ctgtgatcca gaagaaggga agatagagaa ggagggaag 5220  
gatgtaggcg aaggaggtga gagacaggat aggaggaagg aagtggagga ggaggtggta 5280  
ggaattgga ggaggtagaa gccgtgcaga ggaagagggg agagggacga aggaggagcg 5340  
atgaagaaga ggagggagac aaaaaaaggg aag 5373



(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number  
WO 02/008399 A3

(51) International Patent Classification<sup>7</sup>: C12N 15/00,  
15/54, C07K 16/40, C12N 15/63, 9/12, A61K 38/45,  
C12Q 1/48, 1/68

(21) International Application Number: PCT/US01/23092

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/220,038 21 July 2000 (21.07.2000) US  
60/222,112 28 July 2000 (28.07.2000) US  
60/222,831 4 August 2000 (04.08.2000) US  
60/224,729 11 August 2000 (11.08.2000) US

(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(71) Applicant and

(72) Inventor; THORNTON, Michael [US/US]; 9 Medway  
Road, Woodside, CA 94062 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YUE, Henry  
[US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).  
KHAN, Farrah, A. [IN/US]; 3617 Central Road #102,  
Glenview, IL 60025 (US). GURURAJAN, Rajagopal  
[IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US).  
HAFALIA, April, J., A. [US/US]; 2227 Calle de Primav-  
era, Santa Clara, CA 95054 (US). CHAWLA, Naribder  
K. [US/US]; 33 Union Square, #712, Union City, Cal-  
ifornia 94587 (US). ARVIZU, Chandra S. [US/US];  
1706 Morocco Drive, San Jose, California 95125 (US).  
RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird  
Circle, Fremont, CA 94555 (US). GANDHI, Ameena, R.  
[US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025  
(US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis  
Court, San Jose, CA 95118 (US). BAUGHN, Mariah,  
R. [US/US]; 14244 Santiago Road, San Leandro, CA  
94577 (US). TRIBOULEY, Catherine, M. [US/US];  
1121 Tennessee Street, #5, San Francisco, CA 94107 (US).  
BANDMAN, Olga [US/US]; 366 Anna Avenue, MT.  
View, CA 94043 (US). NGUYEN, Danniell, B. [US/US];  
1403 Ridgewood Drive, San Jose, CA 95118 (US). LU,  
Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303  
(US). BURFORD, Neil [US/US]; 105 Wildwood Circle,

Durham, CT 06422 (US). LAL, Preeti [US/US]; P.O. Box  
5142, Santa Clara, CA 95056 (US). DING, Li [CN/US];  
3353 Alma Street #146, Palo Alto, CA 94306 (US). YAO,  
Monique, G. [US/US]; 111 Frederick Court, Mountain  
View, CA 94043 (US). ELLIOTT, Vicki, S. [US/US];  
3770 Polton Place Way, San Jose, CA 95121 (US). RE-  
CIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San  
Francisco, CA 94115 (US). KEARNEY, Liam [IE/US];  
50 Woodside Avenue, San Francisco, CA 94127 (US).  
LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose,  
CA 95123 (US). GREENWALD, Sara, R. [US/US]; 21  
Bucareli Drive, San Francisco, CA 94132 (US). TANG,  
Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA  
95118 (US). XU, Yuming [US/US]; 1739 Walnut Drive,  
Mountain View, CA 94040 (US). WALSH, Roderick,  
T. [IE/GB]; 8 Boundary Court, St. Lawrence Road,  
Canterbury, Kent CT1 3EZ (GB). GIETZEN, Kimberly,  
J. [US/US]; 691 Los Huecos Drive, San Jose, CA 95123  
(US). YANG, Junming [CN/US]; 7125 Bark Lane, San  
Jose, CA 95129 (US). JACKSON, Jennifer, L. [US/US];  
1826 Rina Court, Santa Cruz, California 95062 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
28 August 2003

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.

WO 02/008399 A3

## INTERNATIONAL SEARCH REPORT

International Application No. \_\_\_\_\_

PCT/US 01/23092

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/00 C12N15/54 C07K16/40 C12N15/63 C12N9/12  
A61K38/45 C12Q1/48 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHIER P AND WATT V M: "Primary structure of a putative receptor for a ligand of the insulin family" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, 'Online!' vol. 264, no. 25, 5 September 1989 (1989-09-05), pages 14605-14608, XP002154780 ISSN: 0021-9258 page 14607; figure 1 -& DATABASE EMBL 'Online!' "insulin receptor-related receptor" Database accession no. pl4616 XP002213066  ----- -/-	1-19, 21, 22, 24-45, 65



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

11 September 2002

Date of mailing of the international search report

07.03.03

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Seroz, T

## INTERNATIONAL SEARCH REPORT

Int Application No

PCT/US 01/23092

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>HAENZE J ET AL: "CLONING AND SEQUENCING OF THE COMPLETE CDNA ENCODING THE HUMAN INSULIN RECEPTOR RELATED RECEPTOR" HORMONE AND METABOLIC RESEARCH, THIEME-STRATTON, STUTTGART, DE, vol. 31, no. 2/3, 1999, pages 77-79, XP000944669  ISSN: 0018-5043  page 77, right-hand column, last paragraph  -page 78, left-hand column, paragraph 1  page 78, left-hand column, line 10-13, 32-34</p>	1-19, 21, 22, 24-45, 65
X	<p>WO 00 14212 A (ACTON SUSAN ;MILLENNIUM PHARM INC (US)) 16 March 2000 (2000-03-16)  page 42, line 11 -page 43, line 20  page 56, line 14-17  page 62, line 24-29  page 63, line 28 -page 91, line 16; claims 1-26; examples 1-5</p>	1-19, 21, 22, 24-45, 65
X	<p>SCHULTZ S J ET AL: "IDENTIFICATION OF 21 NOVEL HUMAN PROTEIN KINASES, INCLUDING 3 MEMBERS OF A FAMILY RELATED TO THE CELL CYCLE REGULATOR NIMA OF ASPERGILLUS NIDULANS" CELL GROWTH AND DIFFERENTIATION, THE ASSOCIATION, PHILADELPHIA, PA, US, vol. 4, 1 October 1993 (1993-10-01), pages 821-830, XP000564042  ISSN: 1044-9523  the whole document</p>	1-19, 21, 22, 24-45, 65

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 01/23092**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☒ Claims Nos.: 20, 23  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-19(partially), 21(partially), 22(partially), 24-44(partially), 45(completely), 65(completely)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 45 (completely),  
65 (completely)

Human kinase comprising SEQ ID No 1 and  
a polynucleotide comprising SEQ ID No 21 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

2. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 46 (completely),  
66 (completely)

Human kinase comprising SEQ ID No 2 and  
a polynucleotide comprising SEQ ID No 22 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

3. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 47 (completely),  
67 (completely)

Human kinase comprising SEQ ID No 3 and  
a polynucleotide comprising SEQ ID No 23 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

4. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 48 (completely),  
68 (completely)

Human kinase comprising SEQ ID No 4 and  
a polynucleotide comprising SEQ ID No 24 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

## FURTHER INFORMATION CONTINUED FROM PCT/SA/ 210

5. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 49 (completely),  
69 (completely)

Human kinase comprising SEQ ID No 5 and  
a polynucleotide comprising SEQ ID No 25 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

6. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 50 (completely),  
70 (completely)

Human kinase comprising SEQ ID No 6 and  
a polynucleotide comprising SEQ ID No 26 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

7. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 51 (completely),  
71 (completely)

Human kinase comprising SEQ ID No 7 and  
a polynucleotide comprising SEQ ID No 27 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

8. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 52 (completely),  
72 (completely)

Human kinase comprising SEQ ID No 8 and  
a polynucleotide comprising SEQ ID No 28 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

9. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 53 (completely),  
73 (completely)

Human kinase comprising SEQ ID No 9 and  
a polynucleotide comprising SEQ ID No 29 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

10. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 54 (completely),  
74 (completely)

Human kinase comprising SEQ ID No 10 and  
a polynucleotide comprising SEQ ID No 30 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

11. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 55 (completely),  
75 (completely)

Human kinase comprising SEQ ID No 11 and  
a polynucleotide comprising SEQ ID No 31 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

12. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 56 (completely),  
76 (completely)

Human kinase comprising SEQ ID No 12 and  
a polynucleotide comprising SEQ ID No 32 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

activity of the kinase.

13. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 57 (completely),  
77 (completely)

Human kinase comprising SEQ ID No 13 and  
a polynucleotide comprising SEQ ID No 33 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

14. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 58 (completely),  
78 (completely)

Human kinase comprising SEQ ID No 14 and  
a polynucleotide comprising SEQ ID No 34 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

15. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 59 (completely),  
79 (completely)

Human kinase comprising SEQ ID No 15 and  
a polynucleotide comprising SEQ ID No 35 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of  
PKIN. Method for screening compounds that modulates the  
activity of the kinase.

16. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 60 (completely),  
80 (completely)

Human kinase comprising SEQ ID No 16 and  
a polynucleotide comprising SEQ ID No 36 which encodes and  
identifies said kinase. Expression vectors, host cells,  
antibodies. Methods for diagnosing and treating or  
preventing disorders associated with aberrant expression of



## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

PKIN. Method for screening compounds that modulates the activity of the kinase.

17. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 61 (completely),  
81 (completely)

Human kinase comprising SEQ ID No 17 and a polynucleotide comprising SEQ ID No 37 which encodes and identifies said kinase. Expression vectors, host cells, antibodies. Methods for diagnosing and treating or preventing disorders associated with aberrant expression of PKIN. Method for screening compounds that modulates the activity of the kinase.

18. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 62 (completely),  
82 (completely)

Human kinase comprising SEQ ID No 18 and a polynucleotide comprising SEQ ID No 38 which encodes and identifies said kinase. Expression vectors, host cells, antibodies. Methods for diagnosing and treating or preventing disorders associated with aberrant expression of PKIN. Method for screening compounds that modulates the activity of the kinase.

19. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 63 (completely),  
83 (completely)

Human kinase comprising SEQ ID No 19 and a polynucleotide comprising SEQ ID No 39 which encodes and identifies said kinase. Expression vectors, host cells, antibodies. Methods for diagnosing and treating or preventing disorders associated with aberrant expression of PKIN. Method for screening compounds that modulates the activity of the kinase.

20. Claims: 1-19 (partially), 21 (partially), 22 (partially),  
24-44 (partially), 64 (completely),  
84 (completely)

Human kinase comprising SEQ ID No 20 and a polynucleotide comprising SEQ ID No 40 which encodes and identifies said kinase. Expression vectors, host cells, antibodies. Methods for diagnosing and treating or

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

preventing disorders associated with aberrant expression of PKIN. Method for screening compounds that modulates the activity of the kinase.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## Continuation of Box I.1

Although claims 32, 34 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound.

Although claim 18, 21, 24, are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.

## Continuation of Box I.2

Claims Nos.: 20, 23

Present claims 20, 23 relate to a compound defined by reference to a desirable characteristic or property, namely agonist and antagonist. The claims cover all compounds having this characteristic or property, whereas the application does not provide support within the meaning of Article 6 PCT and disclosure within the meaning of Article 5 PCT. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, no search has been carried out for those claims.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

## INTERNATIONAL SEARCH REPORT

In International Application No.

PCT/US 01/23092

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0014212 A	16-03-2000	US 6183962 B	06-02-2001
		AU 5817799 A	27-03-2000
		CA 2342311 A	16-03-2000
		EP 1112354 A	04-07-2001
		JP 2002524073 T	06-08-2002
		US 6043040 A	28-03-2000
		US 6146841 A	14-11-2000
		US 6180358 B	30-01-2001
		US 6153417 A	28-11-2000
		US 6146832 A	14-11-2000
		US 6190874 B	20-02-2001
		US 6121030 A	19-09-2000
		US 6200770 B	13-03-2001
		US 2002094559 A	18-07-2002
		US 6214597 B	10-04-2001

---

Table 3 (cont.)

SEQ ID NO: ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					Phorbol esters/diacylglycerol binding domain proteins BL00479: H176-G198, H202-C217, L415-L427 Diacylglycerol kinase catalytic domain (presumed) PF00781: K278-K283, P332-F363, R384-L398, C410-Y433, Q441-T461, N772-Y808, L848-G861, V919-Q930 Diacylglycerol/phorbol-ester binding signature PR00008: H202-A213, H214-K226 KINASE PROTEIN DOMAIN PD00584: K74-K84, L386-G395, L466-L473 Phorbol esters/diacylglycerol binding domain: H176-C225	BLIMPS_BLOCKS  BLIMPS_PFAM  BLIMPS_PRINTS  BLIMPS_PRODOM  MOTIFS
12	7170260CD1	268	S161 S188 S255 S29 T15 Y124 Y21		Eukaryotic protein kinase domain pkinae: Y10-L265 Protein kinases signatures and profile protein kinase tyrosine: G82-H162 PROTEIN KINASE DOMAIN DM00004 P27448 58-297: K14-I256 PROTEIN KINASE DOMAIN DM00004 I48609 55-294: K14-S255 PROTEIN KINASE DOMAIN DM00004 Q05512 55-294: K14-S255 PROTEIN KINASE DOMAIN DM00004 JC1446 20-261: Q11-I256 Tyrosine kinase catalytic domain signature PR00109: Y124-L142 Protein kinases ATP-binding region signature: I16-K39	HMMER_PFAM  PROFILES SCAN  BLAST_DOMO  BLAST_DOMO  BLAST_DOMO  BLIMPS_PRINTS  MOTIFS
13	1797506CD1	965	S234 S326 S527 S530 S607 S636 S741 S841 S879 S884 S92 T111 T143 T155 T174 T202 T215 T229 T29 T372 T619 T685 T82 T922 T932 T963 Y173	N227	Eukaryotic protein kinase domain pkinae: F559-F820, Protein kinases signatures and profile protein kinase tyr.prf: E652-G709 PROTEIN KINASE DOMAIN DM00004 Q09499 536-784: P561-A811 PROTEIN KINASE DOMAIN DM00004 P32361 676-970: V564-Q732, T740-A811 KINASE; THREONINE; ATP; SERINE; DM06305 Q09499 786-924: V814-Y949	HMMER_PFAM  PROFILES SCAN  BLAST_DOMO  BLAST_DOMO  BLAST_DOMO